



# **HOMOLOGOUS PROTEINS:**

## Top BLAST Hits:

		Score	E
gb AAF81911.1 AF279265_1	(AF279265) putative anion transpor...	476	e-133
gb AAF71715.1 AF230376_1	(AF230376) prestin [Meriones ungui...	471	e-131
ref NP_000432.1	pendrin [Homo sapiens] >gi 11421915 ref XP...	451	e-125
ref NP_035997.1	Pendred syndrome homolog (human); Pendred'	448	e-124
ref NP_062087.1	Pendred syndrome homolog (human) [Rattus n...	447	e-124
ref NP_067328.1	down-regulated in adenoma [Mus musculus] >...	434	e-120
ref NP_000102.1	down-regulated in adenoma protein [Homo sa...	418	e-115
sp O70531 DTD_RAT	SULFATE TRANSPORTER (DIASTROPHIC DYSPLASI...	365	1e-99
ref NP_000103.1	sulfate anion transporter 1; Diastrophic d...	362	1e-98
ref NP_031911.1	diastrophic dysplasia [Mus musculus] >gi 2...	357	4e-97

## BLAST to dbEST:

	Score	E
gi 8630793 /dataset=dbest /taxon=960...	523	e-146

## **EXPRESSION INFORMATION FOR MODULATORY USE:**

library source:

Expression information from BLAST dbEST hits:

gi|8630793 Human head-neck

Expression information from PCR-based tissue screening panels:

Human fetal lung

```

1 MSQPRPRYVV DRAAYSLTLF DDEFEEKDRT YPVGEKLRNA FRCSSAKIKA
51 VVFGLLPVLS WLPKYKIKDY IIPDLLGGLS GGSIQVPQGM AFALLANLPA
101 VNGLYSSFFP LLTYFFLGGV HQMVPGTFAV ISILVGNICL QLAPESKFQV
151 FNNATNESYV DTAAMEAERL HVSATLACLT AIIQMGLGFM QFGFVAIYLS
201 ESFIRGFMTA AGLQILISVL KYIFGLTIPS YTGPGSIVFT FIDICKNLPH
251 TNIASLIFAL ISGAFLVLVK ELNARYMHKI RFPIPTMIV VVVATAISGG
301 CKMPKKYHMQ IVGEIQRGFP TPVSPVVSQW KDMIGTAFSL AIVSYVINLA
351 MGRTLANKHG YDVDSNQEMI ALGCSNFFGS FFKIHVICCA LSVTLAVDGA
401 GGKSQVASLC VSLVVMITML VLGIIYLYPLP KSVLGALIAV NLKNSLKQLT
451 DPYYLWRKSK LDCCIWVVSF LSSFFLSLPY GVAVGVAFSV LVVVFQTQFR
501 NGYALAQVMD TDIYVNPITY NRAQDIQGIK IITYCSPLYF ANSEIFRQKV
551 IAKTVSLQEL QQDFENAPPT DPNNNQTPAN GTSVSYITFS PDSSSPAQSE
601 PPASAEAPGE PSDMLASVPP FVTFTLILD MSGVSFVDLM GIKALAKLSS
651 TYGKIGVKVF LVNIHAQVYN DISHGGVFED GSLECKHVFP SIHDAVLFAQ
701 ANARDVTPGH NFQGAPGDAE LSLYDSEEDI RSYWDLEQEM FGSMFHAETL
751 TAL

```

# **FEATURES:**

## **Functional domains and key regions:**

[1] PDOC00001 PS00001 ASN\_GLYCOSYLATION  
N-glycosylation site

Number of matches: 3

```

1 153-156 NATN
2 156-159 NESY
3 580-583 NGTS

```

[2] PDOC00005 PS00005 PKC\_PHOSPHO\_SITE  
Protein kinase C phosphorylation site

Number of matches: 2

```

1 45-47 SAK
2 445-447 SLK

```

[3] PDOC00006 PS00006 CK2\_PHOSPHO\_SITE  
Casein kinase II phosphorylation site

Number of matches: 11

```

1 18-21 TLFD
2 158-161 SYVD
3 240-243 TFID
4 365-368 SNQE
5 459-462 SKLD
6 556-559 SLQE
7 635-638 SFVD
8 691-694 SIHD
9 722-725 SLYD
10 726-729 SEED
11 732-735 SYWD

```

[4] PDOC00007 PS00007 TYR\_PHOSPHO\_SITE  
Tyrosine kinase phosphorylation site

Number of matches: 2

```

1 7-15 RYVVDRAAY
2 447-454 KQLTDPYY

```

[5] PDOC00008 PS00008 MYRISTYL  
N-myristoylation site

Number of matches: 10

1	77-82	GGLSGG
2	78-83	GLSGGS
3	89-94	GMAFAL
4	103-108	GLYSSF
5	335-340	GTAFLS
6	435-440	GALIAV
7	481-486	GVAAGV
8	485-490	GVAFSV
9	581-586	GTSVSY
10	681-686	GSLECK

**Membrane spanning structure and domains:**

Helix	Begin	End	Score	Certainty
1	51	71	0.893	Putative
2	82	102	1.020	Certain
3	107	127	1.729	Certain
4	130	150	1.497	Certain
5	186	206	1.723	Certain
6	228	248	1.517	Certain
7	256	276	1.898	Certain
8	288	308	1.252	Certain
9	338	358	1.568	Certain
10	383	403	1.304	Certain
11	412	432	2.345	Certain
12	469	489	1.997	Certain
13	619	639	1.146	Certain

# **BLAST Alignment to Top Hit:**

>gb|AAF81911.1|AF279265\_1 (AF279265) putative anion transporter 1 [Homo sapiens]

Length = 738

Score = 476 bits (1224), Expect = e-133

Identities = 263/724 (36%), Positives = 428/724 (58%), Gaps = 36/724 (4%)

Frame = +3

Query: 54 LFDDEFEEKDR--TYPVGEKLRNAFRCSAKIKAVVFGLLPVLSWLPKYKIKDYIIPDLL 227  
L + EE R + P + R +CS A+ A++ LPVL WLP+Y ++D+++ DLL  
Sbjct: 15 LNQEHLLEELGRWGSAPRTHQWRTWLQCSRARAYALLLQHLPVLVWLP RYPVRDWLLGDLL 74

Query: 228 GGLSGGSIQVPQGMFAFALLANLPAVNGLYSSFFPLLTYYFFLGGVHQMVPGTFAVISILVG 407  
GLS +Q+PQG+A+ALLA LP V GLYSSF+P+ YF G + GTFAV+S++VG  
Sbjct: 75 SGLSVAIMQLPQGLAYALLAGLPPVFGLYSSFYFVFYIFLFGTSRHSISVGTFAVMSVMVG 134

Query: 408 NICLQLAPESKFQVFNATNESYVDTAAMEAERLHVSATLACLTAIIQMGLGFMQFGFVA 587  
++ LAP+ A N+S ++ A +A R+ V++TL+ L + Q+GLG + FGFV  
Sbjct: 135 SVTESLAPQ-----ALNDSMINETARDAARVQVASTLSVLVGLFQVGLGLIHFGFVV 186

Query: 588 IYLSSEFIRGFMTAAGLQILISVLKYIFGLTIPSYTGPGSIVFTFIDICKNLPHTNIASL 767  
YLSE +RG+ TAA +Q+ +S LKY+FGL + S++GP S+++T +++C LP + + ++  
Sbjct: 187 TYLSEPLVRGYTTAAAVQVFSQ LKYVFG LHLSSHSGPLSLIYTVLEVCKWLPQSKVGTV 246

Query: 768 IFALISGAFLVLVKELNARYMHKIRFPIPTMIVVVVATAISGGCKMPKKYHMQIVGEIQ 947  
+ A ++G LV+VK LN + ++ PIP E++ ++ AT IS G + ++ + +VG I  
Sbjct: 247 VTAAVAGVVLVVVKLLNDK LQQQLPMPGELLTLIGATGISYGMGLKHRFEVDVVGNIIP 306

Query: 948 RGFTPTVPSPVVLQWKDMIGTAFSLAIVSYVINLAMGRITLANKHGYDVDSNQEMIALGCSN 1127  
G PV+P + ++G+AF++A+V + I +++G+ A +HGY VDSNQE++ALG SN  
Sbjct: 307 AGLVPPVPAPNTQLFSKLVGSAFTIAVVGFAIAISLGKIFALRHGYRVDSDNQELVALGLSN 366

Query: 1128 FFGSFFKIHVICCALSVTLAVDGAGGKSQVASLCVSLVVMITMLVLGIYLYPLPKSVLGA 1307  
G F+ + C++S +L + GG SQVA SL +++ ++ LG + LPK+VL A  
Sbjct: 367 LIGGIFQCFPVSCSMRSILVQESTGGNSQVAGAISSLFILLIIVKLGE LFDLPKAVLAA 426

Query: 1308 LIAVNLKNSLQKLTDPYYLWRKSKLDCCIWVVSFLSSFFLSLPYGVAVGVAFSVLVVVFQ 1487  
+I VNLK L+QL+D LW+ ++ D IW+V+F ++ L+L G+ V V FS+L+VV +  
Sbjct: 427 IIVVNLKGMRLQSLDMRSLWKANRADLLIWLVTFTATILLNLDLGLVVAVIFSLLL VVVR 486

Query: 1488 TQFRNGYALAQVMDTDIYVNPKTYNRAQDIQGIKIITYCSPLYFANSEIF----- 1637  
TQ + L QV DTDIY + Y+ A++++G+K+ + +YFAN+E +  
Sbjct: 487 TQMPHYSVLGQVPDTDIYRDVAEYSEAKEVRGVKVRSSATVYFANA EFYS DALKQRCGV 546

Query: 1638 -----RQKVIK--TVSLQELQQDFE-NAPPTDPNNNQTPAN-GTSVSYI----- 1760  
++K++ K + L++LQ++ + P N TS+ +  
Sbjct: 547 DVDFLISQKKLLKKQEQLKLQQLKEEKLKQAASPKGASVSINVTSLSDMRSNVED 606

Query: 1761 -----TFSPDSSSPAQSEPPASAEAPGEPDMLASVPPFVTFHTLILDMGVSFVDLMGI 1925  
S D A + ++AP + S + A P FH+LILD+ +SFVD + +  
Sbjct: 607 CKMMQVSSGDKMEDATANGQEDSKAP-DGSTLKALGLPQPDFHSLILDLGALS FVDTVCL 665

Query: 1926 KALAKLSSTYGKIGVKVFLVNIHAQVYNDISHGGVFEDGSLECKHVFP SIHDAVLFAQAN 2105  
K+L + + +I V+V++ H+ V + + G F D S+ KH+F S+HDAV FA +  
Sbjct: 666 KSLKNIFHDFREIEVEVYMAACHSPVVSQLEAGHFF-DASITKKHLFASVHDAVT FALQH 724

Query: 2106 ARDV 2117  
R V  
Sbjct: 725 PRPV 728

FIGURE 2, page 3 of 4

Variable	Mean	SD	Min	Max
Age	34.5	10.2	21	55
Gender	0.5	0.5	0	1
Marital status	0.6	0.5	0	1
Education	12.5	1.5	9	16
Income	15.2	5.8	5	30
Health status	0.8	0.4	0	1
Smoking status	0.3	0.5	0	1
Alcohol consumption	0.2	0.4	0	1
Exercise frequency	0.5	0.5	0	1
Stress level	4.2	1.8	1	7
Sleep quality	3.8	1.5	1	6
Work satisfaction	4.5	1.2	1	6
Life satisfaction	5.2	1.0	1	7
Depression score	2.1	1.5	0	5
Anxiety score	1.8	1.2	0	4
Resilience score	3.5	1.0	1	5
Optimism score	4.0	1.1	1	5
Self-efficacy score	3.2	0.9	1	4
Perceived stress score	2.5	1.3	0	4
Life events score	1.5	0.8	0	3
Support score	3.0	1.0	1	4
Meaning score	3.8	1.2	1	5
Gratitude score	3.5	1.1	1	4
Forgiveness score	3.2	1.0	1	4
Compassion score	3.0	0.9	1	4
Kindness score	2.8	0.8	1	4
Generosity score	2.5	0.7	1	4
Patience score	2.2	0.6	1	4
Humility score	2.0	0.5	1	4
Modesty score	1.8	0.4	1	4
Shame score	1.5	0.3	1	4
Guilt score	1.2	0.2	1	4
Envy score	1.0	0.2	1	4
Jealousy score	0.8	0.2	1	4
Anger score	0.5	0.2	1	4
Dislike score	0.3	0.2	1	4
Disrespect score	0.2	0.2	1	4
Disapproval score	0.1	0.1	1	4
Discomfort score	0.0	0.1	1	4
Displeasure score	0.0	0.1	1	4
Disappointment score	0.0	0.1	1	4
Disillusion score	0.0	0.1	1	4
Disregard score	0.0	0.1	1	4
Disinterest score	0.0	0.1	1	4
Dislike score	0.0	0.1	1	4
Disrespect score	0.0	0.1	1	4
Disapproval score	0.0	0.1	1	4
Discomfort score	0.0	0.1	1	4
Displeasure score	0.0	0.1	1	4
Disappointment score	0.0	0.1	1	4
Disillusion score	0.0	0.1	1	4
Disregard score	0.0	0.1	1	4
Disinterest score	0.0	0.1	1	4

Parsed for domains:

Model	Domain	seg-f	seg-t	hmm-f	hmm-t	score	E-value
PF00916	1/1	187	497 ..	1	328 []	254.5	1.5e-72
PF00189	1/1	651	661 ..	79	89 .]	3.3	8

1	CTGGGTTCCCT	ATGTGGGGAG	GTCATGCTCC	CCACTCATTG	AGCCCCCCCA
51	GGCAAACCAC	CTGGACAGCC	AGACCCATGC	AGACTCTGGA	GCAGGTGGAG
101	AGGAAGAGTG	AGACCACCCC	GCCTCACGGG	CGGTGAAGGG	CCGGCAGCCT
151	CTGAATAGTC	TCTGCTAGGA	GGTAGAAAGC	ACCCCTCCAT	CTTAATCATA
201	GTAATCATCG	CCACTACCAT	TTACTGGGTG	CCTATAAAAG	GCCAGCCTCT
251	TCATACACAT	GATCTCACTG	AATCCTCATA	GCATCTGCCT	GCGAGCTGTA
301	TTATCCCCAT	TTACAGATGA	AGAAACTGAA	CTTTTGAACC	CAGGTCATCT
351	GGCTCTCAAA	TTTGTGCTGT	TTTCCCTAAG	CCACCCGGTC	TCTCATTTCT
401	CCCACTGAAA	TGTCTCACAT	GCCATTGCCC	TTACTCATTT	CTGCCCATGT
451	CTCCTCCAAA	ACACCATTTA	TCAATTGCGT	CAACAAGTAT	GTGTTGAGTA
501	CACACTAAGG	GCCAGGCGAG	GGGCTGGGCA	CAGGCGCTGG	GGGTAGGTTG
551	ATTCTCCCAC	CTTCGCTTGT	GCTGGGTATC	ACCTTGTTGG	CTTTGGCCGG
601	CATCCCCACC	TCACCTGTAG	TTCAAGTGGG	CCTTGGGATC	CCAAGACCAA
651	ATGATGGGAA	TGCACCAGCC	CAGCCTTCAC	CAACTTGAGC	ACAATCTTAT
701	TCATAATAGA	AACTCACATT	TGCATCACAC	TTTACATTTT	ACACAACCCC
751	TTCTTATCCA	TTAACTCATT	TGATCTTCAC	AACAACCTTG	TGAGATATGT
801	CTGTTACTCC	CACTTTAGTG	ATACAGAATC	TGAGGTTTGA	AAAGTAATGC
851	TGACCATTCT	GCCTCATTAA	TAAAGCCAGG	ATTAACCCAG	GCTCCTGGAC
901	CCTTCCACAA	AAGGCATTAA	GGAACCTGCT	CCCTCTGAC	AACTCCCCCT
951	GTCACCCAGG	CTTCTCTCTG	GGAGTTGGG	GGCATCTCTA	GCCCCCAAGT
1001	AGTTACTCAT	TTTCAACCCC	ATCTCAAATC	TTTTGCCAAA	CTGGCCACAG
1051	CCACCCACAC	CTCCCCACCT	CCCAGATACA	AATCCTCACT	CTAAGCCTTC
1101	CCCATCTCTT	TCTTCTCTGT	CCTTCTTTCT	CTGTGGTCTT	CTGAGCAACT
1151	TCTCCCAGTG	CTGGGAGGTA	GAGGGGAGGT	GGGAGACCCA	GTAATTGGAA
1201	GAGGGAGGGG	GAAAGGTTCC	TACAGGGAACT	TCCCTCCGGC	CTCAGGGGCC
1251	CTGGCAGCTA	GCTTGTCCCA	TCTCAGTCC	TGGAACGTCA	GCCAGGTTGC
1301	GCAAAAAGTG	AGGAGGAGAG	GAGCGGCAGT	ACACAAGGGT	GGGGGAAAGA
1351	TTAGGCACAG	GAAGCCGTGG	GAGAGAGAGC	CGGCAGGTGG	ACCATCCTGG
1401	TTTCCCAACA	CACACCATTG	TCCCCCTGGG	AAACCTGTTG	GTGAAGTTCT
1451	AGATGTCTTA	TCCAAGAAGG	GTCCTCTTGA	GGTCATCTCA	GCTATCCCCC
1501	TGCCTCTAGG	CAAGCTGTTT	TCTGTTTCTT	CCAAGCTGAC	TGCTGTAATG
1551	GTAGGAGCCT	TTCTGCCAGT	GAAACTAAGG	TCTGGGAAGG	GAGTATGGCT
1601	TGTGGGGACA	CCAGGGGTCA	GGGGAGGGGA	GGGTCCACCT	GCTGAATCAA
1651	GTGGGGCCTC	CTGCCCTCGT	GATTCCCCCT	TGCCTGGTGC	TCAGTGGGGG
1701	TGATGGTGAC	GCCACAGGTG	TGAGTGCCA	GCCACGTGCT	GAGCGCCAAG
1751	CAAAACAGCC	AGGGTGAGTC	TATGCATCAT	CAGTGCCTGG	GAAGGAAGGC
1801	CACTGCGAGC	AGGGAGTCTG	ACGGAAAAAC	TTGACAGAGG	GAAGGGAGGC
1851	ACCTTGCTTT	ATCGGGCGGG	GGAAGGCCAG	AATAAACTC	TGCTACTGCA
1901	AGGACCAGCT	AGAGAAGGCC	TGGGCTGGCA	CTAGGGAGGG	ATGTTCCCTC
1951	ACCCTCCCCCT	CCTCTGCTTC	TCCCAAAGCT	TGTAATGCC	CCAGATATGA
2001	GCCAGCCCAG	GCCCCGTAC	GTGTTAGACA	GAGCCGCATA	CTCCCTTACC
2051	CTCTTCGACG	ATGAGTTTGA	GAAGAAGGAC	CGGACATACC	CAGTGGGAGA
2101	GAAACTTCGC	AATGCTTTCA	GGTAAGTGGT	CCAGAGCCCC	GACTTCTGCC
2151	TCCTCTGCTC	CTACCAAAAA	TCCTTTCTGC	ACCAGGACAC	GGCTTCTGCA
2201	CTGGTATCCC	TAAGATGGGG	TTAAGGGAAG	CCCTGGGGAA	GTGAGGTTCT
2251	GAATGATGAA	TTTAAGATCC	TACAACCTCA	TCTGTACTGA	GACCCCCAGG
2301	GAGGATGGGG	AGCAGGAGCA	AGAACCATCC	AGAAGGGTTA	TATGGCATTC
2351	CAAACCCCT	GCATGGCATC	TCCCATATTC	TCAATTCAAC	CGGGTCTCTC
2401	TGGGTTTGTT	AAGGTCATGT	AGATGAGCAT	CTACGTTATG	GAGGGGTGGG
2451	GAGCATCAGA	GCCCTTACTC	CATGCCCTGT	TCCCTCCTTA	CAAAAAATAC
2501	CTGAAGTTAC	CATCACCCCA	GGTTCCTTGT	CCTTTCCTTC	CCGGATGTTT
2551	CTTCTCCAC	TTGGTCCAGA	GAATGCCAAA	AGGAGGCCCT	AAATTTCTGA
2601	ACTTTCCTGA	GGGGACCTAC	CAGGGTGTAG	TCCTACCAGG	GCCCAGGGTC
2651	TTCCACTCT	CATCTCCCTG	GAAATGCGAT	GGTGGGTATG	AAACCTTGTC
2701	CCTAAGTAGG	CGCTACACAA	GGTGATCCAT	ACCCACACCC	CAGGAGGCTG
2751	GGGCTGCGGG	TGTCAACCTC	CCCATTTCCA	GACTCCTGGC	AGACCTCCTC
2801	TGCGCCAGCT	ATAGGCCAAC	TCACTCTCCC	TCACTCCCTT	GGGGAAACGG
2851	CTGATTCACT	TACCTGGATT	GAGGTCACAT	GCAATGGCTG	AAGTGGAGAC
2901	GCAGGTGGAA	CTGGTTCAAG	CCGGGGGAAT	CACCCACTTG	AGTTTGTACT
2951	AAAAGCCCCA	GCCCAGCCCT	GTTTCTCTTG	GGAGGCTCCA	TTTCTGCCCA
3001	GTTACAGTCT	GTCCTCACAG	CTGTGCTCCT	CAGACAGGTG	GTCTCTGCCA
3051	GTCTTTGTGC	CAAAGACTTT	AGGGCACAAA	GTCTGAGGAT	GAGAAGATCT
3101	GCTATTGTCC	TAAAGATTA	GGATAATGAA	AGCTGTAAAG	GGATATAGCA

FIGURE 3, page 1 of 20

3151 AACTAACAAT TCCTATGATA CTGGCATGAG AGCCTTGAAC AGTGCCTGGC  
3201 ATAGAGAAGG TGCACCAATA AATATTTGTT TCATGAATGA ATGAATGAAT  
3251 GAATGTCTAG AAAGCTAATC CCTCTCAGCC TCTGTTTCCA GTTCTTCTTT  
3301 CAAGCTTCAG ATTGCTTTGC CCAACATACA GCAGACTTGC AAGTAAGGTT  
3351 GGGCATGGAC TAGCCCTCAA ATGAGTTGTT TTTCTTTCCC TAGCCAGCTC  
3401 TCTATTCTATA AGTCCGGCTT TCTCTGCCAC AAACAGACCT GATGGAGCCC  
3451 CTGCAGGGCT GGTCTCTCTCT TCAAGCAAGG CTTTAGAGTT GCATTAAGCA  
3501 ATTTATCCCC CGTCCACCTC CCCTTCCAGC ATCCCAGGGA TGGCAGAGGC  
3551 ACCCATGAGC CCCAGAAGGG ACAGGGGGTA AGATATTGAT GATGATGCTT  
3601 TTTCTTGGAG TGTTAGTTGG AAGAGAAAAT CTGCCAGAC TTTCCAAGGT  
3651 ACAAGCATT GTCTTTGTTG GTTTCAGTCT TGGGTGACAT CCAGGGGACC  
3701 GAGTGTCAGG GAAACTATTG TTGAGCAAGA GCAAAGAGCA GGAATTGGTG  
3751 CTGGGCAGGA AAGGAAGCCT CATCAGAGCA GGCCAGTGAG TCACCAAATG  
3801 GGCCCTAAGT ATTTGAGTTC CCTCAACTGG GAGAAGGAAA GCAAATGCCC  
3851 CTCACCCACT TCCAGTCATC AATCCACCGG CTGTCACCTT TGAGTTTGTA  
3901 AGCCCTTGTT CCTACCGCTC CTGAGTTTCT ATGAAAGGAC CTTGAGGTGT  
3951 TCAACAAACA AGGAAGGGAT CAACTCTCCC CACCCTGCGT TGACCAATGA  
4001 ATTCTTCCCT CCTCTGCTGC CCAGTGAATT AACAGGAGAA AGAATCCGG  
4051 TATTGGAGTT ACCACACATA AAGGATAGTG AGTCAGCAGA GTGCACCCTG  
4101 CAGGAACAAT AGAGCCTTCC TTTTCAAGGA AGTTCCTAAGA AAAATGGCAG  
4151 CAGGCAGGCC CCACTCGGGT GTATTCACTC ATTCAATTTAT TCAACAAATA  
4201 TTTACTAAGT GCCCCTGTGC AAGGCTCGAG GTGTACAAAG ATGAACAGGA  
4251 GAGCTAGACT TCTTGCCATG CGTGGTGGGG TTTGCTGCCT AGTGGGAGAG  
4301 ACAGACAAA AGCAAGGAAT GCACACACAG GATGCACACA CAGCGGCAGG  
4351 AACCAGGTG CAGTTACCCA GGCTGGGAT CAGACAGACA GGACTCAGAG  
4401 GAGACTTTC CAGAGAAAAG CCATCTGAGC CAAGGGATGG ATCTGATACC  
4451 TCCGAAGGCT GAGCCACCAT AACACTCATA CCTTTAAGCC AAGTCTTATA  
4501 AACTCCCAG GTAAGCAGCT GGCAATCAGA AGACCTCCAG CTAATGCCCA  
4551 GGACAAGTTG ATGAGCTCTC AAGAAAAAGT TCCTGCCTTT TCTTCTCAAT  
4601 ATCCCTGGCA CACAGTTCAG TGAATTTTGA ATGAACCAAT GAATGAAATG  
4651 AGCAGGATAT GATAATCCCT CTCCAACACG GAATGTCCAA GCCATGCAGA  
4701 GCCGACTGGA AATTTTCCCC GTTCCCTTCC AGATGTTCCCT CAGCCAAGAT  
4751 CAAAGCTGTG GTGTTTGGGC TGCTGCCTGT GCTCTCCTGG CTCCCCAAGT  
4801 ACAAGATTAA AGACTACATC ATTCTTGACC TGCTCGGTGG ACTCAGCGGG  
4851 GGATCCATCC AGGTCCCACA AGGTGAAGGG GCTCCTTCAG CCAGGCCTGG  
4901 ATTGCCACTC CCCTCACCAT TCCTCTCCTC ATCCCCACTC CATCCCTCTG  
4951 TGATCCCCAT AAGCTAGTCA TGCTGCTGAG CTTCACTCTC GTTGTCTCTC  
5001 GCAGGCATGG CATTGCTCTC GCTGGCCAAC CTTCTGTCAG TCAATGGCCT  
5051 CTACTCCTCC TTCTTCCCCC TCCTGACCTA CTTCTTCTCT GGGGGTGTTC  
5101 ACCAGATGGT GCCAGGTAAG GCCTCTCCCC TCTGGGCAGG CAGGATGACC  
5151 CAGACCACAA GGATGGGAGG TGTGGCAAAG GGGCCTCGGG AGATTTTCCA  
5201 TCTGCATTCT CCTGGAGTTG TTCTTGGTCA GTCCTAGGGG AATGGTCACT  
5251 GTGAATGTCA TTTCCAGGTC CTCGGTGACC TTGGAGAAAC CACTGAGCCT  
5301 CTTTGAGTTC AGTTAGCATT ACCTGTTCCA TCTTCTCTCT AGGAATGAGA  
5351 GGAAGACTTA GCAGAACAAAG ATATAACCATA TGCTATAACA TGCTTAAACA  
5401 GATGTGAGAA ATCACCATCT AACTCCCTGG TTGGTCCCAG CCGGCCACTA  
5451 CAGGGACATT TGGACTTCTC TGGTGCTAAG TGAGATGGAG GAAAGCCTGG  
5501 TCACAAGGGC TGGTTTCTGG TTCAGGCTCT GCTTATATTT CTTATTTCTG  
5551 AGTTCAATTT CTCACGTGTC CTGTATGACA ATATTGACCA TTGGGGTAAA  
5601 AGCACCTTGA AAAGCATAGA TCATGGTTAG AGTGAGTGGT TGTTATTATT  
5651 GTGTTGGAGA AGAGCCTTGG AGGTGCAGGG ATCCATCCCC CTGGGGTCGG  
5701 GAAGCATTCC TGGGCCCTT TCTGGTTTCC ATCGGTGTGG TTCAAACCTC  
5751 TGATTTTTTG TGGCTGGGTG GGGCACCACA GGTACCTTTG CCGTTATCAG  
5801 CATCCTGGTG GGTAACATCT GTCTGCAGCT GGCCCCAGAG TCGAAATTCC  
5851 AGGTCTTCAA CAATGCCACC AATGAGAGCT ATGTGGACAC AGCAGCCATG  
5901 GAGGCTGAGA GGCTGCACGT GTCAGCTACG CTAGCCTGCC TCACTGCCAT  
5951 CATCCAGGTG AGGGGGCAGC CCCCACCCCT GCTAGAAGGG CATCAGACCA  
6001 CCCTGCCCTT CCCTCAAAGC CTTAGCTTTG ATGCTAAATC TGATTTAGGG  
6051 GGTGTTGGTG GGAGGCTCAT GCCTGTAATC CCAGCACTTT GGGAGGCTGA  
6101 GGAGGGTGGG TCACTTGAGG TCAGGAGTTT GAGACCACCT TGACCAACGT  
6151 GATGAAACCC CATCTCTACC AAAAATACAA AAATAATCCA GGCTTGGTAG  
6201 TATGCGCCTG TAGTCCCACC TACTCAGGAG GCTGAGGCAG GAGAATCACT  
6251 TGAATCCGGG AGGCAGAGGT TGCAGTGAGC TGAGATCGCG CCACTGCACT

FIGURE 3, page 2 of 20



```

6301 CCAGCCTGGG TGACAGAGCG AGACTCCGTC TCAAAAAAAA AAAAAAAA
6351 AAAAAAAA CCCAAGTTAG GGCTCACCTC CTCCCTCCTC CCCATCCCAG
6401 GGCTAAAGTG AACCTTGAAA ATTAACAGTA TCTCCTCATC TGCATGTAGC
6451 AGGACCATAC AAAAAACAA CAGCTGTACC TGGTTAAACT GTCCTGAGCT
6501 TTAAACCTGT AAAAGACTCA CAGCCTCTCT CCATTATCCC GTGGAGAAAC
6551 CCAACTCTCT GCCAGCATAG TCTTGACAGC TGCTAATTTT CTCTAACATC
6601 CCTCACTCCG CTCCAGCCTC CTCTGCTCCA AGCCACAGCA GCAGTTGCAC
6651 AACATAAATT GAGCTTCTGC AAATGGTTGC AAAGGATTCT GCTAGGTTTT
6701 ATGAAGGGAA GCACAACATG ACAGAATGCA AGAGCAAAAC ACAGTCCCAG
6751 AGAGCGCCTT TTCACTCACT CATTCACTCG GTTTTGTGCC AAGAACTAGG
6801 CTAAACCCTG GGATACAAAG ATAAGTAAGA AAGAGGTCCA ATTCACAAGT
6851 TGCTCACAGC CCAGCAGAGG AAGGAGCCAT GTCAACAGAT AAATTGTAT
6901 GCAGTGAGAT AAGCAGCAAA GTAGAGCCAT GTACAAAGAC TGTAGGGACA
6951 CAGAGCAGAG TCACGGAGGA CCTCAAAGAG GAGGTGACAC TCCACCTCTC
7001 TTAAAGGATG AGAACTTAAC CAGGAACAAG GTATACAGAG GATGGTCCAG
7051 GCAGAAGGGA ACAGTGCCTA AAAACACTGA GGCCTGAGAG AGTGTGATCT
7101 GCGCAGGCAA AGTAAGGGGC TTGGTGTGGC TGGAGGGTAG AGGGCCAGA
7151 AGAGGATGGA AAAGTAGGCA GGAGCCAGAC AATGAGATCT GGGGTCTGTT
7201 CTCTGACAGC GACTTTGGGT CTGATTGGCA GTTTATAAGG ATCGTTTGGG
7251 CTACACAATG ATGAGTGGGA GGTGGATTAG AATCAAGGCA GGGGACCTGT
7301 TGGGAGACTC TGACAGAGCC CAGGCAGGAA TAATGCAGGC GAAGACCAGG
7351 TAGAGAAAGA GATGGGGCTG GACTTGAAAA GAATGTTTTA CCAGGAGCTT
7401 GGTGATAGAC TGGATGTGGG AGGTAAGGGA GGATGACTCT CAAGTTTTTG
7451 GTTGGGCAAC CAGTTAATG ATGGTGTCTT TACTGAGAG AGAAAACACT
7501 GGGGGAGGAC TAGACTTATT TTACAGATAA GCCAAAGCCA GAGAGGTGAT
7551 GTGACAGAAA GGCCATGCT CTAAAGGAGC TGAAGGTCTG ATGGCAGCCA
7601 TGTAGAGCAC AGTGAAGGGC AGGTGAAGGT CACAGATGGT CCAATCCCT
7651 CAAGCTACTG CTACGCTAGG ACTGCACGGA GCTCCAGACC TGCCTGTGTG
7701 TGGGGCGGGT CGTTGGAAT GCTGAACCAC ATTGGTCTTC CGCCACCAAC
7751 CACCCTTTTC CTCTCTCAG ATGGGTCTGG GCTTCATGCA GTTTGGCTTT
7801 GTGGCCATCT ACCTCTCCGA GTCCTTCATC CGGGGCTTCA TGACGGCCGC
7851 CGGCCTGCAG ATCCTGATTT CGGTGCTCAA GTACATCTTC GGAAGTACCA
7901 TCCCTCTCTA CACAGGCCA GGTCCATCG TCTTTGTGAG TCTGGGGATG
7951 CACCCTGCC ATTGGAGCAA GGCTCCAGCA GACACATGAG GAGGATGTAC
8001 TGTTTTAAGA TGTCGTGAGC TCCTCATTGC AAGGGCTGGC TTAGCTGTTG
8051 TTCAGAGAGG ATTCTGAGGG GGTTCCTGTC TGGGAGGGT CAAAGTCATG
8101 ACTCACAGAG GTTCTTGTA GTTAATACCT GCAGAAAAGA GCTGTACATT
8151 CTCGCGCAGT TCCCCATTCT AGTGCCTCAA CCCCTCCCTG CCTGGAAAGT
8201 CCTGCCTTAT GTCTAATCTC CATCCCTCCT CCTTCAGCCC AAACCTCTCT
8251 AAAGAAAAAG AAAGCATTC TTTTCTAGCA CAAGTTCCCC ATGTGCCTTT
8301 TGGGAAAGGG CGGTGGGCGA CGGGACAGGG TTCCTGATCA GGGTTTTAAT
8351 TCTGTCTTGG TGTGCCTCCA TTAGCTTTGA TGGCATCCCT TCCCTGGGTC
8401 AGACACCCAA AGGTGGGTA TTATGGGAAG AAGGGGTGGG AGCCTGTGAG
8451 CATGATGCTC TTTCCCCAG ACCTTCATTG ACATTTGCAA AAACCTCCCC
8501 CACACCAACA TCGCCTCGCT CATCTTCGCT CTCATCAGCG GTGCCTTCCT
8551 GGTGCTGGTG AAGGAGCTCA ATGCTCGCTA CATGCACAAG ATTCGCTTCC
8601 CCATCCCTAC AGAGATGATT GTGGTAAGGA CCTTGTTTCA AGCTGGGATG
8651 TTGGGGGGCC AGGCTGTGAG ACGAGGAAGC CCCTACCTTT CCTCACCCCA
8701 TCCCTCAAC TGGCAGCCAG TGGGACAGGA AGTCAGTTGT GAATCCATCC
8751 CATCCCCCGT ATGTGGCGTT TCCTCTCTTT CTAAGTCTCT AATAATTCCC
8801 CTAAGGAGG CAGGGGAGTG GGATTCAGGG TCCCAGAGA AAAGGGAGAC
8851 TTGAGAGAGA CGCCTGCCCT GGCCCCACCT TAGGGCCAAT CCCATTCTC
8901 CACTCTGGGG TTTGCAGGTG GTGGTGGCAA CAGCTATCTC CGGGGGCTGT
8951 AAGATGCCCA AAAAGTATCA CATGCAGATC GTGGGAGAAA TCCAACGCGG
9001 GTGAGTCCAG GTGGCCAGA AGCCTGGCCC ACCCGCACCT CATGCCCCAC
9051 TAAGGCCTGA GCTCGGAGAG GGAGACAAGA TGAAGTCTAT GAAAGTGCAG
9101 TCGAAACTGT ATGACACTGA CCATGTATGA ATTATTACTA TTACCGTTTC
9151 CTGAGAAGGG CCGCACAACC AGCCAATGTA GGCTATTTTA TGAGAAATGA
9201 GTCTTAACTG CCACACTCCC CTTATAAATC TCATTCAACT GATGCTGTTA
9251 AACAAAGCCT CTCTGAACAG CCGCTTGCTG GCTCTTTGCC TTGCTCTAAT
9301 GCATTGGTTC TTTGTCCATG TAGAAAGGGA ACTATTAGGT TCAACCAGAT
9351 TCATGAAGCA TCCACTCTGT GCCAGGCACC ATGCTGGGCC CTGGGAGGAG
9401 AGGGGTGACG CTTGTCTCTG AGGGTTGGAA CAGGCAAGGG AGGGAAGACC

```

FIGURE 3, page 3 of 20

9451 ACATAGCACC AAAGGTCTAG GGGTCTGTGG ACTCGTGAGC ATACAGGGTT  
9501 CAGAATCTGG GAGTTAACAA ACGAGGCCCT ACCACATACT GGCCCGGGGA  
9551 CCTTGGGCAA GTTAGGTTCT CTCAGCCTCA GTTTCCTCCT TTGTAAAACA  
9601 GGAGTGATGG TCCCTACCCT ATGGGGTGGT GCTGAGGATT CAGACTGGAT  
9651 GGGATAACTT AGGCAAAGAT CCCGGCACAC CATGGGGGCC TGGCTGGTCC  
9701 CTGTGGGCTG GTGAAGGACT TGGCTGCCCT CCCCCTCAC ACCCTTGGGT  
9751 TCTGCCCTCT TCTGGCTCC TCGGCAGGTT CCCCACCCCG GTGTCCGCTG  
9801 TGGTCTCACA GTGAAGGAC ATGATAGGCA CAGCCTTCTC CCTAGCCATC  
9851 GTGAGCTACG TCATCAACCT GGCTATGGGC CGGACCCTGG CCAACAAGCA  
9901 CGGCTACGAC GTGGATTCTGA ACCAGGTAGC TCTGGCCACC CCCGGCAGGA  
9951 CTGGGCAGGA CAGGTCAACT CAGGCCTGGC ATGACATATC TTGGGTGGGG  
10001 AGATCATTGG GCTGAGGTGA GGCAGGCTGC CTCGAGTGTG GGGGATAGGG  
10051 GGTCCTCTGA CCCTAAGAGG CTGACCTCCT CTTGACTGGG AATGTGTGAC  
10101 TTTATAGCCA CTGGGTCACT CTCAGGTCTT AGGCCACAG TCCAGCTTGC  
10151 ATGCCCTGACT GCACTTGGTC CCCGTGCCCC CCAGCCCCAC ACTGGCTTCT  
10201 AATCCTGTCT CCTCCCTGCA GGAGATGATC GCTCTCGGCT GCAGCAACTT  
10251 CTTTGGCTCC TTCTTTAAAA TTCATGTCAT TTGCTGTGCG CTTTCTGTCA  
10301 CTCTGGCTGT GGATGGAGCT GGAGGAAAAT CCCAGGTGAG CTTTGTCTTA  
10351 GGGGAGTTGG GGGGAGGTGG TAAGAGAACA GTTGCCCCAA AAAAGCCTGG  
10401 GCACTGCAAG CCAGGCCAGC TCTTCTCCGA CCCCTTCTTC CCGTACTTAG  
10451 TCTCCACTCC ACCAAAGCCA TGGAATTGAA ATAAATCAAG AGCAAAAATT  
10501 TCACACCTTC CCTCTATCCC CAACTCTTTC TCGGAATAGG TGGCCAGCCT  
10551 GTGTGTGTCT CTGGTGGTGA TGATCACCAT GCTGGTCCTG GGGATCTATC  
10601 TGTATCCTCT CCCTAAGGTA AGAGCCCAGC CATCGAGCAG AAGTCAACGA  
10651 AAGACTCCAA TAAGAACAAT CCCTGAGAGT TGTGTGGCAC TTTACGGACC  
10701 ACAAAGTGCC ACTGTTGTCA TACTTAGTCT CAACCACAAA CTGTGAGGTA  
10751 GACAATGCAG GTTTTATCCT CCCATTTTTC CAGGTGAAGG AAAGTGAAGT  
10801 TGAGAGTCTA AGTAACCTTG TCCATAGTGA GGCAGCTTAC AGCGCAGGGC  
10851 TGGTCCCAAA CTCCAGCCTT CTGGCCTCAG AGTCTAATCC CTAGGCAACA  
10901 TTTGCACCTA CCCACGAGTA CCAGGCTCTT ATATAGCCCA GCTAGGAGGG  
10951 CTCTAGGCAT CCGTCAATTA GAGATGAGGG AAGAGAGATA GGGAAAGGAT  
11001 GGGGCCAGGA AGGACCCCAT GGCTCTAACG CCAGCACTTT CCAAACCTAA  
11051 GGTCGAATGC AGAGATTTGG GGGATCAGCC AGGGGAGGTG TTCCAGAAGT  
11101 CCGTCTCTGT CCTGCCAGGC CTTGGGGTCG GGTATGCGCA GGAGGGCAAA  
11151 AAGAAGGGGA GACCCTGGGG TCCTGGAGCA ATGTTCTGCT TCTCTAGTCT  
11201 GTGCTAGGAG CCCTGATCGC TGTCAATCTC AAGAACTCCC TCAAGCAACT  
11251 CACCGACCCC TACTACCTGT GGAGGAAGAG CAAGCTGGAC TGTGTAAGTA  
11301 TCGGGCAGCC TCTGGGTACT GGCCATGCCC CTGCCCTCTC CTCCAACCCC  
11351 ACAGCCCTGT CAGCCCTGTC CTAACAATGA ACCCTCTAGT CTGCTGCTTC  
11401 CTAATTAGCA TGAGATGAGT GGTAAAGT CCGAGTTTCG AAGTGAAACA  
11451 TCCTATGTTT AAACCCCTAAC TCAGCCATCT GCTGGCTCCA TGGCCAATAG  
11501 CAAGCCCCTT AACCTTTCCC AGTCTTGGTG TCTTAACTGG GCAAATGGTT  
11551 ATTTTATGCT CTCTGCCCTC CAGGGTTTTC TATGAAGAAG AAGCAAGGTA  
11601 ATACAAGTAA ACATGTTGTC TACATCGTAT TTTATACTCA ATAAAGCTTA  
11651 GCTATGACTA CTTTATGACA TACAGCTTTA AAAAACAATA GGAAATAGTT  
11701 TGTATTTTAA AAAAAACCT AGAACATAAA GCCAGAGGAC CAAAATCTTG  
11751 AGCAAGTTAC TAGACTTCCC TGGGGTCTA TTTCTCATC TGTAAATGGG  
11801 GGTGAGACTC ATGCAGTCAT GGTGCGTCA AACGCTGGTT CCGAGGATTA  
11851 AATGAGATCC CAGTGGGAAA ACACCGCATG AGCGCAAACA TTCTGCAAAC  
11901 ATGACTTATT GTCCTGATTA GTCACACACT CCACCGCATC ATCCGCTGGG  
11951 CATAGTAATG AAGGCCAGTG TGTTTTGACG AACTGCCTT CTCTCCATTT  
12001 AAGCCCCACC ATAACCTATG GGAGAGGATT TACTAACTT TCTTAACGGT  
12051 GATGAAACCA AGGCTCAGAA TGGTTAAGTA AATTGTCAAA GGCCACAGAG  
12101 GTAGGGAGTG GTAGAGTCTG GATTAAAACT CCAAGTCTTG GACTCCAGAC  
12151 CTCTAGGCTG TACTGTCTCA TAGGGAAGGC AGTCTCACC ACCTAGGGCA  
12201 GAGAAGAAAA TCCTTAAAGC CAGAGAAGTG AGTGGTCTAT CTGTGGTCAC  
12251 CCAGAGAGAC AGTGATGAGG ACAGGGAGAA AAATTATACC TCAGTTCCCA  
12301 GCCCAAGGAT CTGCTTTGAC CATAACCCAA CAAGCCCCCG CTATGGTGGT  
12351 ATTTCTTTAG GTTCATATGG CGGCTTTTGT TTCCATTGTA TCTTCACAGC  
12401 AATTCTCTAC AGGAATCTGG GCAGATTTAT TTCCTTTAGA GGAATTTCCA  
12451 GGTCTTAAAA TCTATAGGGG GCAACTATCA AAATTTACC CAATGTTGCC  
12501 CCCTACCCAC ACACAAAACC AGGCCCCCAG CCGATCAGAA AGCACTGCTG  
12551 AGCTCCTGTC AGGGCCCCAG CAGCTCGCTG TGAGACAGAG AGAGGGAAGT

FIGURE 3, page 4 of 20

12601 CACATTTATT GATCACCTAC TGAGCATCCA TCACTAGGCT AGGACCGTCA  
12651 CATTCCTTAA CTTTTGAATC CTTTCATGAG GTAGGCATTA TTATTCTCCT  
12701 TTTGTTTTAC ATAGCCATTA AAGAACAAAA TTTGGGGCTG GGTGTGCTGA  
12751 CTCACACCTG TGATCTAGCA CTTTAGGGGG CTGAGGCAGG AGGATCGCTT  
12801 GAAGTCAGGA TTTCAAGGTC AGCTTGGGCA GCTTAGCGAG AGCCGTCTCT  
12851 AGAAAAATAT AAAAGTTAGC TGGGTGTGGT GGCACGTGCC TATAGTCCTA  
12901 ACTATTCAGG AAGGTTAGGC GGGAGCACAA CTTGGGTTC AGGGTTTGAG  
12951 GCTCCAGTGA GCTGATCTTG CCACTGCACT ACAGCCTGAG CAACAGAGCA  
13001 AGACCCTGTG ACTCCAAAAA CAAACAAACA AACACATTTT GAACCCAAAC  
13051 AGATCTGACC CAAGATGCAT GCTCTTATAG ATGCCACCTC CCTGTGTGCT  
13101 GGGGCTTCTA CTA AAAACAC AGACAAGATC AGGCAACCAC AGTCAATCTA  
13151 AGGGAAAGAG GAAAGTGTAA CCAAAGCACA AATACATAAA ATATTGCAAA  
13201 AATGCTATTT AAAGAAAAAA AAGAGAAGAG AGGCTCTGAG GTTGTACTAA  
13251 CAGAGAATGG CTTTGGCTAA TCCAGGAAGA CTTCCTGAAA GAGGTTGTTT  
13301 TTTCCCCAGG TCTGCTTTTG ACATCTCTCT TTTACAGTG CATCTGGGTA  
13351 GTGAGCTTCC TCTCCTCCTT CTTCTCAGC CTGCCCTATG GTGTGGCAGT  
13401 GGGTGTGCGC TCTCCGTCC TGGTCGTGGT CTTCCAGACT CAGTTGTAAG  
13451 TGATAGCTTC CGCCCTCCTA GGCCACAGT CGGTTCCCTG GGCCAGCCCCG  
13501 CAAAGGGCTT CCATGCCACG GCCTGGCTTA GTCCACTGTA CCTTCCACCT  
13551 CTGGGCCTGG CACTGGAGGT GCTGCCAGGC CCAAAGAGAG CCCAACCCAG  
13601 CCAGGACTGT GGGCACAGTC TGGGCTGTTT GACTTCCCAT ATCTTGAAAA  
13651 CCCAGAGAA AGCCAGCATA CTCTTGCTGG GGATGGCTGG GGAGAGGGCA  
13701 GTGGCAGAGA AAGGAGGGCA AGGGCAGGTG GTGAGATTCA ACATCCTTCC  
13751 AAAGACATTG CCAGAACCCC AAACCAATG GGACCCACC CCAGGAGAGC  
13801 GCCAGGGTGG AAGACAGAAG CTGTGTTCTA CACACTGGGA GTATTACAGA  
13851 GAAGGGGTCT TGGCCAAGGC AGGGAGTACG CTGAATGTTG GGGGAATCCT  
13901 ATCTTCTCTT CTTGAGAACT CAGAACAAGG AAATGATGAC TTCAGGGCGA  
13951 CTCCCACCAC TTCTCCCACC ACTTCTCTCC CCTGCCCTGT GGTCTGGGAG  
14001 CTATGTCAAG GACCTGCCTG TCATCCTCAT AGTTATAGGA GGCCACAGGC  
14051 CACCAGACAT GTGTCTCCAG TGCAAAAAGA CAGACACAGC AAGTCTGGGG  
14101 GTGAGGACAG GACCCCATCC TACCTTGGCT CTGCCCCCGC CCCAGCAGGG  
14151 GCACCCCTCC AGGCCCATGT GCCATTAGCA TTCTCTTATG TTTTCTCTT  
14201 CCTGCTTCAT CCAGTCGAAA TGGCTATGCA CTGGCCAGG TCATGGACAC  
14251 TGACATTAT GTGAATCCCA AGACCTATAA TAGGGTAGGT AATTCAAGCT  
14301 TATGACCTCC TTCTTTTGCT CTGCACCACC CCAAGAAGAG GTTGCTTTTT  
14351 AAAGCCAATA AAGACATTTT TGCAACTTGA GCTCAGTCTC CCTGTACAG  
14401 GCCCAGGATA TCCAGGGGAT TAAAATCATC ACGTACTGCT CCCCTCTCTA  
14451 CTTTGCCAAC TCAGAGATCT TCAGGCAAAA GGTATCGCC AAGGTAAGGC  
14501 TCAGTCCCTG GCGACCAGAG GCTCTGGACA GAGAGTGGCC GGAAAATGGA  
14551 AGCAGAAGGG CGGTGGGAGC TGAGAATAGG CCACTCCCAT AGAGGGTGGA  
14601 GGTCAAGATT GCTGTTGGCT CTCTCCCTGC AGACAGGCAT GGACCCCCAG  
14651 AAAGTATTAC TAGCCAAGCA AAAATACCTC AAGAAGCAGG AGAAGCGGAG  
14701 AATGAGGCCC ACACAACAGA GGAGGTCTCT ATTATGAAA ACCAAGGTGA  
14751 ATGAAGGCCA GAAGCAGCCC CGTGCCTGC TCTCCTGCCC ATTCTGATAC  
14801 TGCCCCCTGT TACTCATGGT ACCCTGGGGG CCCCCTTCC CACCCTGACA  
14851 GGCAAAGACA GAAAGTCTCT GGGAACTAG CTTGGTGGCC GCTGGGCATT  
14901 TTTCTTCTTT TTTTCTTTT TCTTTTAGA GATGGAATTT TGCTCTTGTC  
14951 ACCCAGGCTT GAGTGCAATG GCGTTATCTT GGCTCACTGC AACCTCCACC  
15001 TCTGGGGTTC AAGCGATTCT CCTGCCTTAG CCTCCCAAGT CGCTGAGATT  
15051 ACAGGTGCCA CCACACCAG CTAATTTTG TATTTTATG AGATATTGGG  
15101 TTTCACCATG TTGGCCAGGC TGGTGTCAAA CTCCTGACCT CAGGTGATCC  
15151 ACCTACCTTA GCCTTCCAAA GTGCTGGGAT TACAAGCCTG AGCCACTGCG  
15201 CCCAGCCTGG GCATTTTTCT TCTTGGATGA GGTGCTACCA TCTCCAGGG  
15251 AAGCCACTGA ACCCCCAAGG CCCTTCTCCA TTTTCTGGCT AAGATAGGAC  
15301 ATGGCCCATG GACTTTTGAA CAACCCAGAG GGGGAACAGC AGTGAATTTT  
15351 CTGGGGAACC CAGGCAGCCC AGGGCTAGCA AGGCTGGGGT GGCCATGGCA  
15401 GTAATCCTTG TAATCCCAGC ACTTTAGGAG GCCGAGATGG GAGAATCACT  
15451 CTCATGAGTT CAGGAGTTCG AGACCAGCCT GCCCAACGTG GCGAAACGCT  
15501 GTCTCTACTA AAAATACACA AAAATTAGCC AGGCGTGGTG GTGGGCACCT  
15551 GTAATCCCAG CTAATCAGGA GGCTGAGGCA CGAGAATCAC TTGAACCCGG  
15601 GAGGCAGAGG TTGCAAGTGA CCGAGATAGT GCCACTGCAC TCCAGCCTAG  
15651 GCAACAGAGG GAGACTCTGT CTCAAGAAAT AAAGGAGCTC AGTGTCCCCG  
15701 GAGGGGCTTT CTCCCAGAGA GAGTGGGCTT GAGGCTTCAG TGCCTCTCTT

FIGURE 3, page 5 of 20

15751 GGCTGGGTCC TCTGACTTTG TCTGGGTTGT AGGAGACCAA GTTTCAGGC  
15801 CCTGCCTAAG AAAGGGCTTT GGGAGAGGCC TCTCTGGTGG AGCTTTCAGG  
15851 GTCTGTGTTT ACCATCACCG AGGCGAGTTA TTCCCTTACA CCTACACCCT  
15901 CCATGCCCTT GCTTCAGTCA CAGCAAGGTC TGGCTCAGTC TGGTGGTCCC  
15951 TGA CTCTGCC CACTGTCCCC ACCCTTCCAG ACTGTCTCCC TGCAGGAGCT  
16001 GCAGCAGGAC TTTGAGAATG CGCCCCCAC CGACCCCAAC AACACCAGA  
16051 CCCC GGCTAA CGGCACCAGC GTGTCTTATA TCACCTTCAG CCTGACAGC  
16101 TCCTCACCTG CCCAGAGTGA GCCACCAGCC TCCGCTGAGG CCCCCGGCGA  
16151 GCCCAGTGAC ATGCTGGCCA GCGTCCCACC CTTTCGTACC TTCCACACCC  
16201 TCATCCTGGA CATGAGTGGA GTCAGCTTCG TGGACTTGAT GGGCATCAAG  
16251 GCCCTGGCCA AGGTGAGGCC CTCGGGGACA GCAAGCACCA CCCACTCCAC  
16301 CCCCTCCGCT CTGCTCTCCA CATTCCCTTT CCTGGGAGCC CTCATTTTCA  
16351 GAAGCTGAGG GAGGAAGCTC ACTGGGGAGA CTAACAGCTC CTAGGAATCC  
16401 CTCCTTTCCC CAGACGCCAC CAGGTTGAGA CATTCTCCAC AGAGCAGGCC  
16451 CAGACGGCCC ATGACAATGA GTGGCGGGAC AAGTCTACCA GAGTTTCAGG  
16501 CCCCTGTGCT CCCAACACCC CCAGCAGTGG CCATCCCAAG TCCCTCTCAG  
16551 CCATCAGGAA CCCACCCAGG TTCTCTGAGG AGGGTCCAGT TTGGCTCCTG  
16601 GTTCATGATC TGCTGCCCTT GTCCCTCATT CACCAGCCAC CCTAGGACAG  
16651 GAGAAGAAAT AATACCAGTG CCCACACCA TCAGGCCAAA CAGAGAGCCC  
16701 ACGGGACACC TTGAATGAAT GTATCCATCT GATAACTTTC CAGCAGCCAC  
16751 CGCCATGGC GGGAGTCAGC AAACCTCAGA GCTGGCTCAG ATAGAGGCAA  
16801 GCCAGGGGAA CAATGGGCAC AGAGAGTGT CCGACTGCCT TCACCATCAA  
16851 CCAGGCGCAG GGCAGGCCCC ATACCCAGCC TTGGGCTCA GCCGGCTTCC  
16901 TTAGCCAGGA TCTGGAGTCC AGGCCAGCCT TGGCTGAAGC TCTAGACTCC  
16951 CTGAGCCTCC ATCCTCCCCT GCAGCTTCTG TCTGAAGCCA CAAAGAAGTC  
17001 TGAGAATCTA AGCTACTGAA AGAAAAGATC AGCCGGGCGT GGTGGCTCAC  
17051 TCCTGTAATC CCAGCACTTT GGGAGGCCAA GGCAGGTGGA TCACAAGGTC  
17101 AGGAGTTCAA GACCAGCCTG GCCAACATGG TGAAACCCCG CCTCTACTAA  
17151 AAATACAAAA ATTAGCCAGG TGTGGTGACG GGCCCTGTA GTCCCAGCTA  
17201 CTCGGTAGGC TGAGGCAGAG AATTGCTTGA ACCCAGGAGG CGGAGGTTGC  
17251 AGTGAGCCAA GATCGGCCA CTGCACTCCA GCCTGGGCAA CAGAGTGAAA  
17301 CTCCATCTCA AAAGAAAAA AAAGAAAATA TCTAGCCCCA CAAGAAGGGG  
17351 CCATGGTGAC TTTAAGTGCC CGCCACGTTG GCAAAAGTCC ATTTCCGCTC  
17401 CACTTCCCAG AGAAACCGTC AGCCAACACT CCAGGGAGAA GTGGTGTGCT  
17451 TTGCTGCTAT TTTTGTCTTT GGCTGCTGGG CTCTCAGGGT TGCTTATTTG  
17501 TTTGGCTTCC CCTCTGAAGT ACGTTTTGTG AATCACTTTT GAGACCCACT  
17551 CAGAACATTC CTTTCTTTT GCCTCCCTAC CCCAACACA CTTCTAGCTG  
17601 AGCTCCACCT ATGGGAAGAT CGGCGTGAAG GTCTTCTTGG TGAACATCCA  
17651 TGGTAAGAGA AAGAGGACAT TTAGGGACTG AAAGACTGGC AAGGAGTGTG  
17701 GGGTAGGAAC AGGTTGTTGG GGTCTGAATA GTGAGGAGGT TGGAAACGAG  
17751 AGCACCCAGC TATCCCCCAC AAGCTGCTGC CTGCTCATAA AAGCTTCAGG  
17801 TACAAGTCCA AAGAGACTGG TCAGATTGCA TAAACATCCT AGGGGCCTTA  
17851 GTGACAGAGT GGGGGTGAGG AGGTCAATGA GTTACAGAAG GACAGCTAGG  
17901 ATTCTAATCT ACCCCATAAC TAATTGCGA CGTATCCTTG GCCGAGTCAC  
17951 TTTATCTCTC AAGGGATCTA TTTCTACCTA TGTAAAACGA GAGGGTTGAC  
18001 TAGATGGATT TGGGGATCCT CTCCCAATCA GAAACTCTGT GAATCGATAT  
18051 AGGCATAGAG CACACGGTAC CCTAATTTCC CAGGGAACAT ATAAATATGC  
18101 AGTTTTGTAG GCATACAGCC TCCAAAGGGT GCATATACAC AGCCTCAAGG  
18151 ACGTGGCCAC AGGGCAGCAG ACATTTACAT GACTAGCATG TACGCAAAGT  
18201 GCAGAGATGT GGGAGCAAAGT GCACACAGAC ACACAGGAGA ATGTGAAGGG  
18251 GCACATACAC ACACACCCAG CTCCCTGCAC TGGGTGAGC CCCCTCCAGC  
18301 AGGGCTGCAG TTCCCAAGCT CCGCATGGCC ACGTTCGGGG AGAGAATCTG  
18351 CAGTGGCAAT GACCTGCTAT GATATGTTCT GGAGTTAGAA GCAGTGGATT  
18401 CTCCTCAACC TCACTGGACA CCCCCTTAGG AAACCATCTC TAGGATTAAG  
18451 AGTAATCCAC ACAAACTTCC AATGCCACAC ATTGGAAGTT GCTGGAAGG  
18501 TCTGGGAAAA CAAGAGGAAG GATGGGTCTT TGGGGGATAG AACTGGCAGC  
18551 GGCCTCTTCA AGGATGGCTT AGGCTTTTCC ACTCGAATCA CCACAAAGTA  
18601 CTGACTCCCT AAATCAAAC TCTTCTTCT GCTCTGGGT GAACTTCAG  
18651 CATCTCAAG TTCATGTTGC CCTCTGCCGT CCAGAACTGA TATTGCACTG  
18701 CCAATGCCAT GCCCTCAGA TACAGCAAGA GCTGGGACCT CAGGCTCTC  
18751 CCATCCCTGC TCTGGTCTCA CTATCTTCCC CACCCCGAGC TCCAATCCAC  
18801 AATGGCTGTT ATCTTTCTGA AGGTGATCTT TTCTCCTTCT AGCCAGGTG  
18851 TACAATGACA TTAGCCATGG AGGCGTCTT GAGGATGGGA GTCTAGAATG

FIGURE 3, page 6 of 20

18901 CAAGCACGTC TTTCCCAGCA TACATGACGC AGTCCTCTTT GCCCAGGCAA  
18951 ATGCTAGAGA CGTGACCCCA GGACACAAC TCCAAGGGGT AAGGTTCTTG  
19001 CACCTGGGGA ATCCTAGGCT CCAAGGCACT GAAATAGCAG GACCAAGAGG  
19051 CATTATTAGA AAGAACACAG GAGAAGGTTT AAGTTCCAAT ATCAAGTCTG  
19101 CCATTTTCAGT TTTCTGAATC TGTTCCTTA TCTATAGAAT GAGCACCATC  
19151 AACTAACATT ACCTACCTCT CTGCATTTTT CTTTTATTTT GTTTTAGGGT  
19201 TAAATGATAA TTACATCTTT TGTGTCACTT GAAAGCACTT TGTGTATTGT  
19251 AAAAATTCTT TATCAATATA AGTTTTCTGG TTGCACAAAC ACCCAAAGCA  
19301 TAGTAGAGCA GGCCCACTCT GCTGGCATCG TTCCCTGCCT CCTCCTCATC  
19351 TCTTTCTAAA GGGGGCTTTC GGAAGGGAG GGGAGGGGAG TAAGCCTACC  
19401 CATTTTAACT TACCGGAGCT TAGAGATTTT AGGCTGGTGA GGGATAAAGA  
19451 GATTGGGTCT GAGTTTTGTC TCAGCTTTTT GACATTTAAT TTAGTAGCTC  
19501 AGTAAGTCAT ACAAATGGGA TACAAATAAC ACCATCTAAA ACTCCAGAAG  
19551 ACTGGGGAGT CAGAAAAATC CTACCTCCTT GGGGTCCCTG CCCAGATCCC  
19601 CAGTCATCTC TAGCCCTCAG GGTCCCCTCC CAGCTCAGCT CCTGCCCTTG  
19651 GCCTCCCAAG ACTCTTGTTG TGCCCCAGCC CTGGGTAAAA ACCTCCCCTG  
19701 CCCTCTGTGG GTCATAAGAA AGGCTTTTCT GGCCTTAGAG CAATGATTTG  
19751 CTCTTTGCCT TAAGAGACTG ATGAAGGTGA AACCATCTGT TCTAAGTGCT  
19801 GAAAGACTGC CCAGGAACAC ACAGGGCGCT GGCTCCTGCC CTCCATGCCT  
19851 AGAGGGAAAC CCTGGGGAAA CAACGGGCTT TCCTGCTTCG TGAAATTGT  
19901 CCGCAGAGCA AAGAGGGAGA TTCTGGAGGA AGCTGCATTA GTTGTAGTG  
19951 CCCTAATCAT GTTCAGCTAC TCTAGTTGGT ATGTATACTT GATTAGTCAT  
20001 AGCACTTATA AATAATTTAT ATTTTATATA ATATATACTT ACATATTATA  
20051 GACCAATCAC AGATACAAAT CACACACATA AACACACACC TTTTCAACAG  
20101 CATTGTGAGG GACAAAGCAG GCAAAGTGAG GCTGGTTATC AGACTTTAAC  
20151 AGATTAGAAA ATATATTCCC AGGAGGACAG GAATTCCCCA AGGTCAAGCA  
20201 GCTAGCCAAT AGTTTTTCTA AGCTGAGTAA AACCTTCCCT GCCTCTAACG  
20251 GCCCACAAG GAGGGAAGAC CGCGATACAC ACCTGTCTGG TATAAGGGGG  
20301 AAGACCACAG CCGTGCTGTT TTTGTGAGGC AGGTAAGGGA AGGGGCAAGA  
20351 GGATAAGTCA TGTGTCAGGA AGCAGCGTCC AACCAGAGCC GGCCACCTGT  
20401 CCCTTTTCTT GCCACCATGC ACCAACTTTG CTGTTCACTC ACTGAAGCTC  
20451 ATTCTGCACT GGCTTCCTCC CTTCAGGCT CCAGGGGATG CTGAGCTCTC  
20501 CTTGTACGAC TCAGAGGAGG ACATTGCGAG CTACTGGGAC TTAGAGCAGG  
20551 TGAGCTGAGG GAAGGGGCTG TGAGGGTGGG AGCAGGGCGA AGAGGGGAAG  
20601 GATGGGGTCG CTGTCAAATA CAAGGCGTTC ACTCAGCTGT CTCACCTCCA  
20651 GCCCAGAGCA GTCACATTCA AGGCCACAAA GATTGTGGT CATCTTTGTT  
20701 TTTTTTCTTT TCCTTTTCTT TTTTTTTTTT TTTTAATTTG AGACAAAGTC  
20751 TCACTCTATC ACCCAGACTG GAATGCAGTG GCATGATCTC AGCTCACTGC  
20801 AACCTCTGCC TCCCGGGTTC CAGAGGTTCT CCTGCCTCAG CCTCCGAGT  
20851 AGCTGGGACT TCAGGCCTGC GCCCAGCTAA TTTTGTATT TTAGTAGAG  
20901 ACAGCTTTTC ACCATGTTGG CTGGGCTGGT CTCGAACCTC CGATCTCAAG  
20951 CAATCTGCCT GCCTCGGTCT CTAAGTGCC TGGATTACAG GCATAAGCCA  
21001 CGATGCCTGG CCTTTGTTTT CATTCTTCTC ACTCCCTGAA AGGCATCGTG  
21051 GGGAGAGGGT GAGTCACTGG ACCAAGTCCT AGAGAACCAG TATCTATTCT  
21101 TATTCTCCAA CACATCACCC ACGTGACCCT GAGCAAGCCA CATAACCCT  
21151 GGGCCCTAGT TTTTATCATC TGTGAAATTA GGGGAAACAT AGGTAATACC  
21201 TGTCCCATCC ACCACACAAG ATTGGCAGGG CAGTCACTTG TTCTTTCAAT  
21251 AATTCAGCAG GTATTTATGG CGTACCTACT GTTGCCTGA CACAGTTCAG  
21301 GATGGGCACA TAGCAGTGAG CAAAACAAAG GCCTCTGCCT TTTAGAAACT  
21351 TACGTTATGG TAGAATAGAT GGATTTNNNN NNNNNNNNNN NNNNNNNNNN  
21401 NNNNNNNNNN NNNNNNNNNN NNNNNNGTCT ACAAATGAAT TATTATTGCA  
21451 TGTGGACAAG CCTTAAGAAC TAAAAAATAT GTGGCTGGGT GCAATGGTTC  
21501 ACACCTGTAA TCCAGCACT TTGGGAGGCT GAGGTGGGCG GACCACCTGA  
21551 GGTGAGGAGT TTGAGACCAG CCTGGCCAAC ATGGCGAAAC CCCGTCTCTA  
21601 CTAAGACAC AAAAATTAGC CAGGCGTAGT GGTGCATGCC TGTAGTCCCA  
21651 GCTACTCGGA AGTCTGAGGC ATGAGAATCA CTTGAACCTG GGAGGCAGAT  
21701 GTTGCACTGA GCCGAGATCG TGCCACTGCA CTCCAGCTTG GGTGACAGAG  
21751 CTAGACTGTC TCAAAAACAA ACAAACAAAA CAAAACCTAA AAGATATGTG  
21801 GATATGAGGG ATCACCATCC CCATAGGGCC CCTGGATTAA CACCACCCCA  
21851 CCAATGCCCT GAATTAAGAG AAACCAGATG ACTAGGTTTG GAGAAATCTG  
21901 GCTTTGGGTC TATGAGAAGT AGTGTCTCTC TTTGTGCCTC TTCCCATCT  
21951 TTTTGACATT GAGCTCCATG GTGCTCTGAA TCCGTCTCTC ACAGTGCTGA  
22001 TGGCAGGTGG GACAGATTAG AAAATAGAGC TGGAGCCACA GAGATTGGC

FIGURE 3, page 7 of 20

22051	AGACTGATTT	CGGTGCCCTC	TTGGAATCTC	CAGCACATTC	CAAAAAGCCT
22101	GGATAGGACC	AAAATAGCTT	ATCAACGTGA	GAAAGGACTT	CAGAGCTTGT
22151	CTACTGCCAA	CCCTCATTTT	ACCCATGAG	GAAAGTGAAG	CTATTAGGGG
22201	GCAGGGGACA	CGTGGAAGGT	CACACAGCAC	ACAGGAGGTG	ATTACATGT
22251	AGATTTACAG	ACCTGCTCCT	GCCACGCTGG	ACTGGTTCAC	CTCCTAGGCT
22301	GACCCGTGCT	TCCCCCTGCT	CACACACAGT	CCTGCCACACA	CACACACACA
22351	CACACACACA	CACAGGTGCT	TTGTTCTGGC	CAGGGGTTCC	TAGGGTCACC
22401	TCTTGTTTGC	AGCCACTGTG	ACCCCAACTG	GTCTAACCTC	TCTCTTCCCC
22451	TCCCACTTCC	TTCTGTGGT	TCCTGCAGGA	GATGTTCTGG	AGCATGTTTT
22501	ACGCAGAGAC	CCTGACCGCC	CTGTGAGGGC	TCAGCCAGTC	CTCATGCTGC
22551	CTACAGAGTG	CCTGGCACTT	GGGACTTCCA	TAAAGGATGA	GCCTGGGGTG
22601	ACAGGGGGTG	TCGGGCGGAG	GAAAGTGTCAT	CCCCAGAGC	TTGGGTTCTCT
22651	CTCTCTCTCT	CCCTCTCTCT	CTCCCTTCTCT	TCCCTCCCCG	CATCTCCAGA
22701	GAGAGCCTCT	CAGCAGCAGG	GGGGTGCTAC	CCTTACAGGA	GTGAGAGTCT
22751	GGTGAGCCCA	CTCTTCACCC	GTGAGGCCCT	GGCCGCAATG	GACAAGCCTC
22801	CTGCTCACTC	CACCCACCC	ACCTCTGCCC	TGTCCTTGGC	AGCTGAAGGA
22851	CACCTTGACT	TCCAGCTTTT	ACGAGTGAGC	CAAAAACAGA	AGGACAAGTA
22901	CAACTGTGCT	GGCCTGCTGT	ACAAGCTTCA	AAAAGTGTC	CAGAGCCCTA
22951	ACGGCTCGGT	GTAGATGGT	GTGAGGCTGT	CACGGACATA	GGGATAAACT
23001	TGGTTAGGAC	TCTGGCTTGC	CTTCCCCAGC	TGCCCTCACT	CTGTCTCTGG
23051	CAGCTCTGCA	CCCAGGGACC	ATGTGCTCTC	CACACCCAGG	AGTCTAGGCC
23101	TTGGTAACTA	TGCGCCCCCC	GTCCATCATC	CCCAAGGCTG	CCCAAACCAC
23151	CACTGCTGTC	AGCAAGCACA	TCAGACTCTA	GCCTGGACAG	TGGCCAGGAC
23201	CGTCGAGACC	ACCAGAGTCA	CCTCCCCGGG	GACAGCCAC	TAAAGTTCTG
23251	CCTCAGCCTC	CTGAAACATC	ACTGCCCTGA	GAGGCTGCTC	CCTTCCCCTG
23301	GAGGCTGGCT	AGAAAACCCA	AAGAGGGGGA	TGGGTAGCTG	GCAGAATCAT
23351	CTGGCATCCT	AGTAATAGAT	ACCAGTTATT	CTGCACAAAA	CTTTTGGGAA
23401	TTCCTCTTTG	CACCCAGAGA	CTCAGAGGGG	AAGAGGGTGC	TAGTACCAAC
23451	ACAGGGAAAA	CGGATGGGAC	CTGGGCCCAG	ACAGTCCCCC	TTGACCCACG
23501	GGCCCATCAG	GGAATAGCCT	CCCTTTGGTA	AATCTGCCTT	ATCCTTCTTT
23551	ACCTGGCAAA	GAGCCAAATC	TGTTAACTCT	TCCTTTACAG	CCTTGGCCCC
23601	AGAGACACAA	TGGGGTCTTT	CTTAGGCCAA	AGGTGGAAGT	CCTCCAGGGA
23651	TCCGCTACAT	CCCCTAACTG	CATGCAGATG	TGGAAGGGGG	CTGATCCAGA
23701	TTGGGTCTTC	CTGCACAGGA	AGACTCTTTA	ACACCCCTAG	GACCTCAGGC
23751	CATCTTCTCC	TATGAAGATG	AAAATAGGGG	TTAAGTTTTC	CATATGTACA
23801	AGGAGGTATT	GAGAGGAACC	CTACTGTTGA	CTTGAAATTA	AATAGGTTCC
23851	ATGTGTAACT	GTTTTGTAAA	ATTTAGCTGT	AAATGCACAG	AAATCTTCT
23901	GGCCTCTCAT	CACCTGTTTT	CTCAAGCTTC	TTGAGTTTAA	CAACCCCTTC
23951	CCTAACAGGT	TGGGCTGGCC	CAGCCTAGGA	AAACATCCCC	ATTTCTAACT
24001	TCAGCCAGAC	CTGCGTTGTG	TGTCTGTGTG	TTGAGTGAGC	TGGTCAGCTA
24051	ACAAGTCTTC	TTAGAGTTAA	AGGAGGGGGT	GCTGGCCAA	AGCCAACACA
24101	TTCTTGGCCC	AGGAGCAATT	CTTTTCTGTG	AATTCATTAT	GCCATCTGGC
24151	TGCCAATGGA	ACTCAAAACT	TGGAAGGCTG	AGGACAATGT	TATCTGGGAT
24201	TACCGTGCA	CAGCACCCGA	AGTGCCAAAT	TCCAGGAGGA	CAAGAGCCTT
24251	AGCCAATGAC	AACTCACTCT	CCCCTACTCC	ACCTCCTTCC	AAGTCCAGCT
24301	CAGGCCCAGG	AGGTGGGAGA	AGGTCACAGA	GCCTCAGGAA	TTTCCAAGTC
24351	AGAGTCCCCT	TTGAACCAAG	TATCTAGATC	CCCTGAGGAC	TTGATGAAGT
24401	GATCCTTAA	CCCCAAGTAA	TCATTAACCC	CCAGACCAGC	CTCAGAAGTG
24451	AAGGAGATTG	TTGACCCAGT	GACCTGGAGT	TGAGGCTCAG	GGAGAGATCT
24501	GCCACATGTC	TGAGGGTTCG	AGAGCC		

### FEATURES:

Start: 1997  
Exon: 1997-2121  
Intron: 2122-4732  
Exon: 4733-4872  
Intron: 4873-5004  
Exon: 5005-5115  
Intron: 5116-5781  
Exon: 5782-5957  
Intron: 5958-7770  
Exon: 7771-7935  
Intron: 7936-8470

Exon: 8471-8623  
 Intron: 8624-8917  
 Exon: 8918-9000  
 Intron: 9001-9777  
 Exon: 9778-9925  
 Intron: 9926-10221  
 Exon: 10222-10335  
 Intron: 10336-10539  
 Exon: 10540-10617  
 Intron: 10618-11197  
 Exon: 11198-11293  
 Intron: 11294-13338  
 Exon: 13339-13445  
 Intron: 13446-14214  
 Exon: 14215-14284  
 Intron: 14285-14400  
 Exon: 14401-14493  
 Intron: 14494-15980  
 Exon: 15981-16262  
 Intron: 16263-17597  
 Exon: 17598-17652  
 Intron: 17653-18842  
 Exon: 18843-18988  
 Intron: 18989-20477  
 Exon: 20478-20549  
 Intron: 20550-22478  
 Exon: 22479-22523  
 Stop: 22524

**CHROMOSOME MAP POSITION:**

Chromosome 1

**ALLELIC VARIANTS (SNPs):**

DNA				Protein		
Position	Major	Minor	Domain	Position	Major	Minor
48	C	G	Beyond ORF(5')			
132	G	A	Beyond ORF(5')			
724	A	C	Beyond ORF(5')			
1558	C	G	Beyond ORF(5')			
1577	A	G	Beyond ORF(5')			
2487	C	A	Intron			
2634	T	C	Intron			
4352	A	G	Intron			
5157	A	C	Intron			
5658	A	T	Intron			
5945	T	C	Exon	180	T	T
6281	C	T	Intron			
6452	G	C	Intron			
6610	T	G	Intron			
7247	T	C	Intron			
7360	A	G	Intron			
7644	A	T	Intron			
8127	A	C	Intron			
8317	G	A	Intron			
9079	G	A	Intron			
9537	G	T	Intron			
12302	C	G	Intron			
12354	C	T	Intron			
12487	C	T	Intron			
13198	-	A	Intron			
13257	A	G	Intron			
14541	G	A	Intron			

14545	A	G	Intron			
15041	C	A	Intron			
15053	A	C	Intron			
15065	A	G	Intron			
15108	A	C	Intron			
16274	-	G	Intron			
17424	C	T	Intron			
17627	G	A	Exon	657	V	V
18427	T	C	Intron			
18813	C	G	Intron			
19035	T	C	Intron			
19182	T	C	Intron			
19508	-	G C	Intron			
19571	T	G C	Intron			
20147	T	G	Intron			
20180	G	A	Intron			
20584	A	T	Intron			
20717	T	C	Intron			
20894	A	G	Intron			
21787	-	A C	Intron			
22264	T	C	Intron			
22338	-	C A	Intron			
23363	T	C	Beyond ORF (3')			
23688	G	A	Beyond ORF (3')			
24210	A	C	Beyond ORF (3')			

Context:

DNA

Position

48	CTGGGTTTCCTATGTGGGGAGGTCATGCTCCCCACTCATTGAGCCCCC [C,G] CAGGCAAACCACCTGGACAGCCAGACCCATGCAGACTCTGGAGCAGGTGGAGAGGAAGAG TGAGACCACCCCGCCTCACGGGCGGTGAAGGGCCGGCAGCCTCTGAATAGTCTCTGCTAG GAGGTAGAAAGCACCCCTCCCATCTTAATCATAGTAATCATCGCCACTACCATTACTGGG TGCCTATAAAAGGCCAGCCTCTTCATACACATGATCTCACTGAATCCTCATAGCATCTGC CTGCGACTGTTATTATCCCATTACAGATGAAGAACTGAATCTTTGAACCCAGGTCAT
132	CTGGGTTTCCTATGTGGGGAGGTCATGCTCCCCACTCATTGAGCCCCCCCAGGCAAACCAC CTGGACAGCCAGACCCATGCAGACTCTGGAGCAGGTGGAGAGGAAGAGTGAGACCACCCC GCCTCACGGGC [G,A] GTGAAGGGCCGGCAGCCTCTGAATAGTCTCTGCTAGGAGGTAGAAAGCACCCCTCCCATCT TAATCATAGTAATCATCGCCACTACCATTACTGGGTGCCTATAAAAGGCCAGCCTCTTC ATACACATGATCTCACTGAATCCTCATAGCATCTGCCTGCGACTGTTATTATCCCATT ACAGATGAAGAACTGAATCTTTGAACCCAGGTCATCTGGCTCTCAAACCTTGTGCTGTTT TCCCTAAGCCACCCGGTCTCTCATTTCTCCCACTGAAATGTCTCACATGCCATTGCCCTT
724	ATTGCCCTTACTCATTTCTGCCCATGTCTCCTCCAAAACACCATTTATCAATTGCTCAA CAAGTATGTGTTGAGTACACACTAAGGGCCAGGCGAGGGGCTGGGCACAGGCGCTGGGGG TAGGTTCAATCTCCACCTTCGCTTCTGCTGGGTATCACCTGTGGGGTCTTGCCGGGCAT CCCACCTCACCTGTAGTTCAAGTGGACCTTGGGATCCAAGACCAATGAATGGAATGC ACCAGCCAGCCTTCACCAACTTGAGCACAATCTTATTATAATAGAACTCACATTTGC [A,C] TCACACTTTACATTTTACACAACCCCTTCTTATCCATTAACCTCATTTGATCTTCACAACA ACCCTGTGAGATATGTCTGTACTCCCACTTTAGTGATACAGAATCTGAGGTTTGAAAAG TAATGCTGACCATTTCTGCCTCATTAATAAAAGCAGGATTAACCCAGGCTCCTGGACCCCT CCACAAAAGGCATTAAGCAACCTGCTCCCTCTGACAACCTCCCCTGTCACCCAGGCTCT CCTCTGGGAAGTTGGGGGCATCTTAGCCCCAAGTAGTTACTCATTTTCAACCCATCT
1558	TCAGCTCTGCCCATCTCAGCTCCTGGAACGTGAGCCAGGTTGCGCAAAAAGTGAGGAGGA GAGGAGCGGCAGTACACAAGGTGGGGGAAAGATTAGGCACAGGAAGCCGTGGGAGAGAG AGCCGCGAGGTGGACCATCTGTTTCCCACACACACCATTTGCCCTGGGAAACCTG

FIGURE 3, page 10 of 20



TTGGTGAAGTTCTAGATGTCTTATCCAAGAAGGGTCCTCTTGAGGTCTCTCAGCTATCC  
CCCTGCCCTCTAGGCAAGCTGTTTTCTGTTTCTTCCAAGCTGACTGGCTGAATGGTAGGAG  
[C, G]  
CTTTCTGCCAGGAACTAAGGTCTGGGAAGGGAGTATGGCTTGTGGGGACACCAGGGGT  
CAGGGGAGGGGAGGGTCCACCTGCTGAATCAAGTGGGGCCTCTGCCCTCGTGATTCCCC  
TTTGCTTGGTGCTCAGTGGGGGTGATGGTGACGCCACAGGTGTGGAGTGCCAGCCACGTG  
CTGAGCGCCAAGCAAAACAGCCAGGGTGAGTCTATGCATCATCAGTGCCTGGGAAGGAAG  
GCCACTGCGAGCAGGGAGTCTGACGGAAAACTTGACAGAGGGAAGGGAGGCACCTTGCT

1577 CTCTGGAACGTGAGCCAGGTTGCGCAAAAAGTGAGGAGGAGAGGAGCGGCAGTACACAA  
GGGTGGGGGAAAGATTAGGCACAGGAAGCCGTGGGAGAGAGAGCCGGCAGGTGGACCATC  
CTGGTTTCCCCACACACACCATTTGCCCCCTGGGAAACCTGTTGGTGAAGTTCTAGATGT  
CTTATCCAAGAAGGGTCTCTTGAGGTCTCTCAGCTATCCCCCTGCCTCTAGGCAAGCT  
GTTTTCTGTTTCTTCCAAGCTGACTGGCTGAATGGTAGGAGCCTTTCTGCCAGGGAACT  
[A, G]  
AGGTCTGGGAAGGGAGTATGGCTTGTGGGGACACCAGGGGTGAGGGGAGGGGAGGGTCCA  
CCTGCTGAATCAAGTGGGGCCTCTGCCCTCGTGATTCCCCCTTGCTGGTGTCTCAGTGG  
GGGTGATGGTGACGCCACAGGTGTGGAGTGCCAGCCACGTGCTGAGCGCCAAGCAAAACA  
GCCAGGGTGAGTCTATGCATCATCAGTGCCTGGGAAGGAAGGCCACTGCGAGCAGGGAGT  
CTGACGGAAAACTTGACAGAGGGAAGGGAGGCACCTTGCTTTATCGGGGCGGGGAAGGC

2487 ACACGGCTTCTGCACTGGTATCCCTAAGATGGGGTTAAGGGAAGCCCTGGGGAAGTGAGG  
TTCTGAATGATGAATTAAGATCCTACAACCTCATCTGTACTGAGACCCCCAGGGAGGAT  
GGGAGCAGGAGCAAGAACCATCCAGAAGGGTTATATGGCATTCCCAAACCCCTGCATGG  
CATCTCCCATATTCTCAATTACCCGGGTCTCTCTGGGTTTGTAAAGGCATGGTAGATGA  
GCATCTACGTTATGGAGGGGTGGGGAGCATCAGAGCCCTTACTCCATGCCCTGTTCCCTC  
[C, A]  
TTACAAAAAATACCTGAAGTTACCATCACCCAGGTTCTTTGTCCTTTCCCTCCCGGATG  
TTCCTTCCCTCACTTGGTCCAGAGAATGCCAAAAGGAGGCCCTAAATTTCTGAACTTTCC  
TGAGGGGACCTACCAGGGTGATAGTCTACCAGCGCCAGGGTCTTTCCACTCTCATCTCC  
CTGGAAATGCGATGGTGGGTATGAAACCTTGTCCCTAAGTAGGCGCTACACAAGGTGATC  
CATACCCACACCCAGGAGGCTGGGGTGTGCGGGTGTACCCTCCCCATTCCCAGACTCCT

2634 AGGGTTATATGGCATTCCCAAACCCCTGCATGGCATCTCCCATATTCTCAATTCACCCGG  
GTCTCTCTGGGTTTGTAAAGGCATGGTAGATGAGCATCTACGTTATGGAGGGGTGGGGAG  
CATCAGAGCCCTTACTCCATGCCCTGTTCCCTCCTTACAAAAAATACCTGAAGTTACCAT  
CACCCAGGTTCTTTGTCTTTCCCTCCCGGATGTTCCCTCCTCACTTGGTCCAGAGAA  
TGCCAAAAGGAGGCCCTAAATTTCTGAACTTTCTGAGGGGACCTACCAGGGTGATGCC  
[T, C]  
ACCAGCGCCAGGGTCTTTCCACTCTCATCTCCCTGGAAATGCGATGGTGGGTATGAAAC  
CTTGTCCCTAAGTAGGCGCTACACAAGGTGATCCATACCCACACCCAGGAGGTGGGGC  
TGCGGGTGTACCCCTCCCCATTCCCAGACTCCTGGCAGACCTCCTCTGGCCAGCTATAG  
GCCAACTCACTCTCCCTCACTCCCTTGGGGAACGGCTGATTCACTTACCTGGATTGAGG  
TCACTGGCAATGGCTGAAGTGAGACGCGAGGTGGAAGTGGTTCAGGCCGGGGGAATCACC

4352 ATTGGAGTTACCACACATAAAGGATAGTGAGTCAGCAGAGTGCACCCTGCAGGAACAATA  
GAGCCTTCCTTTTCAAGGAAGTTCTAAGAAAAATGGCAGCAGGCAGGCCCCACTCGGGTG  
TATTCATCTATTCAATTTATTCAACAAATATTTACTAAGTGCCCTGTGCAAGGCTCGAGG  
TGTACAAAGATGAACAGGAGAGCTAGACTTCTTGCCATGCGTGGTGGGGTTTGCTGCCTA  
GTGGGAGAGACAGACAAAAAGCAAGGAATGCACACACAGGATGCACACACAGCGGCAGGA  
[A, G]  
CCAAGGTGCAGTTACCCAGGCTGGGATCAGACAGACAGGACTCAGAGGAGACTTTCCCA  
GAGAAAAGCCATCTGAGCCAAGGGATGGATCTGATACCTCCGAAGGCTGAGCCACCATAA  
CACTCATACCTTTAAGCCAAGTCTTATAAACTCCCCAGGTAAGCAGCTGGCAGTCAGAAG  
ACCTCCAGCTAATGCCAGGACAAGTTGATGAGCTCTCAAGAAAAAGTTCTTGCCTTTTC  
TTCTCAATATCCCTGGCACACAGTTTCACTGAATTTTGAATGAACCAATGAATGAAATGAG

5157 ATCCAGGTCCCACAAGGTGAAGGGGCTCCTTCAGCCAGGCCTGGATTGCCACTCCCCTCA  
CCATTCTCTCTCATCCCCACTCCATCCCTCTGTGATCCCCATAAGCTAGTCATGCTGC  
TGAGCTTCAGTCTCGTTGTCTCTGACGGCATGGCATTGCTCTGCTGGCCAACCTTCCT  
GCAGTCAATGGCCTCTACTCCTCCTTCTTCCCCCTCCTGACCTACTTCTTCTGGGGGGT  
GTTACCAGATGGTGCCAGGTAAGGCCTCTCCCCTCTGGGCAGGCAGGATGACCCAGACC  
[A, C]

FIGURE 3, page 11 of 20

CAAGGATGGGAGGTGTGGCAAAGGGGCCTCGGGAGATTTTCCATCTGCATTCTCCTGGAG  
TTGTTCTCTGGTCAGTCCTAGGGGAATGGTCACTGTGAATGTCATTTCCAGGTCCTCGGTG  
ACCTTGGAGAAACCACTGAGCCTCTTTGAGTTCAGTTAGCATTACCTGTTCCATCTTCCT  
CCTAGGAATGAGAGGAAGACTTAGCAGAACAAGATATACCATATGCTATAACATGCTTAA  
ACAGATGTGAGAAATCACCATCTAACTCCCTGGTTGGTCCCAGCCGGCCACTACAGGGAC

5658 TTAGCAGAACAAGATATACCATATGCTATAACATGCTTAAACAGATGTGAGAAATCACCA  
TCTAACTCCCTGGTTGGTCCCAGCCGGCCACTACAGGGACATTTGGACTTCTCTGGTGC  
AAGTGAGATGGAGGAAAGCCTGGTCAACAAGGGCTGGTTTCTGGTTCAGGCTCTGCTTATA  
TTTCTTATTTCTGAGTTCATTTTCTCACGTGTCCTGTATGACAATATTGACCATTGGGGT  
AAAAGCACCTTGAAAAGCATAGATCATGGTTAGAGTGAGTGGTTGTTATTATTGTGTGG  
[A, T]  
GAAGAGCCTTGGAGGTGCAGGGATCCATCCCCCTGGGGTCGGGAAGCATTCCTGGGCCCC  
TTTCTGGTTTCCATCGGTGTGGTTCAAACCTCTGATTTTGTGGCTGGGTGGGGCACCA  
CAGGTACCTTTGCCGTTATCAGCATCCTGGTGGGTAACATCTGTCTGCAGCTGGCCCCAG  
AGTCGAAATTCCAGGTCTTCAACAATGCCACCAATGAGAGCTATGTGGACACAGCAGCCA  
TGGAGGCTGAGAGGCTGCACGTGTGAGCTACGCTAGCCTGCCTCACTGCCATCATCCAGG

5945 ATTATTGTGTTGGAGAAGAGCCTTGGAGGTGCAGGGATCCATCCCCCTGGGGTCGGGAAG  
CATTCCTGGGCCCCCTTCTGGTTTCCATCGGTGTGGTTCAAACCTCTGATTTTGTGGC  
TGGGTGGGGCACCAAGGTACCTTTGCCGTTATCAGCATCCTGGTGGGTAACATCTGTCT  
GCAGCTGGCCCCAGAGTCGAAATTCAGGTCTTCAACAATGCCACCAATGAGAGCTATGT  
GGACACAGCAGCCATGGAGGCTGAGAGGCTGCACGTGTGAGCTACGCTAGCCTGCCTCAC  
[T, C]  
GCCATCATCCAGGTGAGGGGGCAGCCCCAACCCCTGCTAGAAGGGCATCAGACCACCCTG  
CCCCCTCCCTCAAAGCCTTAGCTTTGATGCTAAATCTGATTTAGGGGGCTGGGTGTGGAGG  
CTCATGCCTGTAATCCAGCACTTTGGGAGGCTGAGGAGGGTGGATCACTTGAGGTGAGG  
AGTTTGAGACCACCTTGACCAACGTGATGAAACCCCATCTCTACCAAAAATACAAAATA  
ATCCAGGCTTGGTAGTATGCGCCTGTAGTCCCACCTACTCAGGAGGCTGAGGCAGGAGAA

6281 GCTAGAAGGGCATCAGACCACCCTGCCCTCCCTCAAAGCCTTAGCTTTGATGCTAAATC  
TGATTTAGGGGGCTGGGTGTGGAGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCTGA  
GGAGGGTGGATCACTTGAGGTGAGGAGTTTGGAGCACCTTGACCAACGTGATGAAACCC  
CATCTCTACCAAAAATACAAAATAATCCAGGCTTGGTAGTATGCGCCTGTAGTCCCACC  
TACTCAGGAGGCTGAGGCAGGAGAATCACTTGAATCCGGGAGGCAGAGGTTGCAGTGAGC  
[C, T]  
GAGATCGCGCCACTGCACTCCAGCCTGGGTGACAGAGCGAGACTCCGTCTCAAAAAAAAAA  
AAAAAAAAAAAAAAAAAAAAACCAAGTTAGGGCTCACCTCCTCCCTCCTCCCCATCCCAGG  
GCTAAAGTGAACCTTGAAAATTAACAGTATCTCCTCATCTGCATGTAGCAGGACCATAACA  
AAAAAACACAGCTGTACCTGGTTAAACTGTCTGAGCTTTAAACCTGTAAAAGACTCAC  
AGCCTCTCTCCATTATCCCGTGGAGAAACCAACTCTCTGCCAGCATAGTCTTGCACT

6452 ATGAAACCCCATCTCTACCAAAAATACAAAATAATCCAGGCTTGGTAGTATGCGCCTGT  
AGTCCCACCTACTCAGGAGGCTGAGGCAGGAGAATCACTTGAATCCGGGAGGCAGAGGTT  
GCAGTGAGCTGAGATCGCGCCACTGCACTCCAGCCTGGGTGACAGAGCGAGACTCCGTCT  
CAAAAAAAAAAAAAAAAAAAAAAAAAAACCAGTTAGGGCTCACCTCCTCCCTCCTCC  
CCATCCCAGGCTAAAGTGAACCTTGAAAATTAACAGTATCTCCTCATCTGCATGTAGCA  
[G, C]  
GACCATACAAAAAACACAGCTGTACCTGGTTAAACTGTCTGAGCTTTAAACCTGTAA  
AAGACTCACAGCCTCTCTCCATTATCCCGTGGAGAAACCAACTCTCTGCCAGCATAGTC  
TTGCAGACTGCTAATTTTCTCTAACATCCCTCACTCCGCTCCAGCCTCCTCTGCTCCAAG  
CCACAGCAGCAGTTGCACAACATAAATTGAGCTTCTGCAAATGGTTGCAAAGGATTCTGC  
TAGGTTTTATGAAGGAAGCACAAATGACAGAATGCAAGAGCAAAACACAGTCCCAGAG

6610 GTGACAGAGCGAGACTCCGTCTCAAAAAAAAAAAAAAAAAAAAAAAAAAACCAGTTA  
GGGCTCACCTCCTCCCTCCTCCCATCCAGGGCTAAAGTGAACCTTGAAAATTAACAGT  
ATCTCCTCATCTGCATGTAGCAGGACCATACAAAAAACAACAGCTGTACCTGGTTAAAC  
TGTCCTGAGCTTTAAACCTGTAAAAGACTCACAGCCTCTCTCCATTATCCCGTGGAGAAA  
CCCAACTCTCTGCCAGCATAGTCTTGCACTGCTAATTTTCTCTAACATCCCTCACTCC  
[T, G]  
CTCCAGCCTCCTCTGCTCCAAGCCACAGCAGCAGTTGCACAACATAAATTGAGCTTCTGC  
AAATGGTTGCAAAGGATTCTGCTAGGTTTTATGAAGGAAGCACAAATGACAGAATGCA  
AGAGCAAAACAGTCCCAGAGAGCGCCTTTTCATTCACTCATTTCGGTTTTGTGCC

FIGURE 3, page 12 of 20

AAGAACTAGGCTAAACCTGGGATACAAAGATAAGTAAGAAAGAGGTCCAATTACAAAGT  
TGCTCACAGCCCAGCAGAGGAAGGAGCCATGTCAACAGATAAATTTGTATGCAGTGAGAT

7247 GACACAGAGCAGAGTCACGGAGGACCTCAAAGAGGAGGTGACACTCCACCTCTCTTAAAG  
GATGAGAACTTAACCAGGAACAAGGTATACAGAGGATGGTCCAGGCAGAAGGGAACAGTG  
CCTAAAAACACTGAGGCCTGAGAGAGTGTGATCTGCGCAGGCAAAGTAAGGGGCTTGGTG  
TGGCTGGAGGGTAGAGGGCCCAGAAGAGGATGGAAAAGTAGGCAGGAGCCAGACAATGAG  
ATCTGGGGTCTGTTCTCTGACAGCGACTTTGGGTCTGATTGGCAGTTTATAAGGATCGTT  
[T, C]  
GGGCTACACAATGATGAGTGGGAGGTGGATTAGAATCAAGGCAGGGGACCTGTTGGGAGA  
CTCTGCAGAGGCCAGGCAGGAATAATGCAGGCGAAGACCAGGTAGAGAAAGAGATGGGG  
CTGGACTTGAAAAGAATGTTTTACCAGGAGCTTGGTGATAGACTGGATGTGGGAGGTAAG  
GGAGGATGACTCTCAAGTTTTTGGTTGGGCAACCAGGTTAATGATGGTGTATTTACTGA  
GAGAGAAAACACTGGGGGAGGACTAGACTTATTTTACAGATAAGCCAAAGCCAGAGAGGT

7360 AACAGTGCCTAAAAACACTGAGGCCTGAGAGAGTGTGATCTGCGCAGGCAAAGTAAGGGG  
CTTGGTGTGGCTGGAGGGTAGAGGGCCCAGAAGAGGATGGAAAAGTAGGCAGGAGCCAGA  
CAATGAGATCTGGGGTCTGTTCTCTGACAGCGACTTTGGGTCTGATTGGCAGTTTATAAG  
GATCGTTTGGGCTACACAATGATGAGTGGGAGGTGGATTAGAATCAAGGCAGGGGACCTG  
TTGGGAGACTCTGCAGAGGCCAGGCAGGAATAATGCAGGCGAAGACCAGGTAGAGAAAG  
[A, G]  
GATGGGGCTGGACTTGAAAAGAATGTTTTACCAGGAGCTTGGTGATAGACTGGATGTGGG  
AGGTAAAGGAGGATGACTCTCAAGTTTTTGGTTGGGCAACCAGGTTAATGATGGTGTGAT  
TTACTGAGAGAGAAAACACTGGGGGAGGACTAGACTTATTTTACAGATAAGCCAAAGCCA  
GAGAGGTGATGTGACAGAAAGGCCCATGCTCTAAAGGAGCTGAAGGTCTGATGGCAGCCA  
TGTAGAGCACAGTGAAGGGCAGGTGAAGGTCACAGATGGTCCAATTCCCTCAAGCTACTG

7644 GACCAGGTAGAGAAAGAGATGGGGCTGGACTTGAAAAGAATGTTTTACCAGGAGCTTGGT  
GATAGACTGGATGTGGGAGGTAAGGGAGGATGACTCTCAAGTTTTTGGTTGGGCAACCAG  
GTTAATGATGGTGTGATTTACTGAGAGAGAAAACACTGGGGGAGGACTAGACTTATTTTA  
CAGATAAGCCAAAGCCAGAGAGGTGATGTGACAGAAAGGCCCATGCTCTAAAGGAGCTGA  
AGGTCTGATGGCAGCCATGTAGAGCACAGTGAAGGGCAGGTGAAGGTCACAGATGGTCCA  
[A, T]  
TTCCCTCAAGCTACTGCTACGCTAGGACTGCACGGAGCTCCAGACCTGCGTGTGTGGG  
GCGGGTCTGTGGAAGTCTGAACACATTGGTCTTCCGCCACCAACCACCTTTTCCTCC  
TCTCAGATGGGTCTGGGCTTCATGCAGTTTGGCTTGTGGCCATCTACCTCTCCGAGTCC  
TTCATCCGGGGCTTCATGACGGCGCGGCTGCAGATCCTGATTTTCGGTGTCTCAAGTAC  
ATCTTCGGACTGACCATCCCCTCCTACACAGGCCAGGGTCCATCGTCTTTGTGAGTCTG

8127 CATCCGGGGCTTCATGACGGCCGCCGCTGCAGATCCTGATTTTCGGTGTCTCAAGTACAT  
CTTCGGACTGACCATCCCCTCCTACACAGGCCAGGGTCCATCGTCTTTGTGAGTCTGGG  
GATGCACCCCTGCCATTGGAGCAAGGCTCCAGCAGACACATGAGGAGGATGTACTGTTTT  
AAGATGTCGTGAGCTCCTCATTGCAAGGGCTGGCTTAGCTGTTGTTTACAGAGAGGATTCTG  
AGGGGGTTTCTGTCTTGGGAGGGTCAAAGTCATGACTCACAGAGGTTCTTGGTAGTTAAT  
[A, C]  
CCTGCAGAAAAGAGCTGTACATTCTCCGCCAGTTCCCCATTCTAGTGCCTCAACCCCTCC  
CTGCCTGGAAAAGTCTGCCTTATGTCTAATCTCCATCCCTCCTCCTTCAGCCCAAACCTCT  
TCTAAAGAAAAGAAAGCATTCTTTTCTAGCACAAAGTTCCCCATGTGCCTTTTGGGAAA  
GGGCGGTGGGCGACGGGACAGGGTTCTGATCAGGGTTTAAATTCTGTCTTGGTGTGCCT  
CCATTAGCTTTGATGGCATCCCTTCCCTGGGTGAGACACCCAAAGGTGGGGTATTATGGG

8317 GAGCTCCTCATTGCAAGGGCTGGCTTAGCTGTTGTTTACAGAGGATTCTGAGGGGGTTTC  
TGTCTTGGGAGGGTCAAAGTCATGACTCACAGAGGTTCTTGGTAGTTAATACCTGCAGAA  
AAGAGCTGTACATTCTCCGCCAGTTCCCCATTCTAGTGCCTCAACCCCTCCCTGCCTGGA  
AAGTCTGCCTTATGTCTAATCTCCATCCCTCCTCCTTCAGCCCAAACCTCTTCTAAAGAA  
AAAGAAAGCATTCTTTTTCTAGCACAAAGTTCCCCATGTGCCTTTTGGGAAAGGGCGGTGG  
[G, A]  
CGACGGGACAGGGTTCTGATCAGGGTTTAAATTCTGTCTTGGTGTGCCTCCATTAGCTT  
TGATGGCATCCCTTCCCTGGGTGAGACACCCAAAGGTGGGGTATTATGGGAAGAAGGGGT  
GGGAGCCTGTGAGCATGATGCTCTTTCCCCCAGACCTTCATTGACATTTGCAAAAACCTC  
CCCCACACCAACATCGCCTCGCTCATCTTCGCTCTCATCAGCGGTGCCTTCCCTGGTGTG  
GTGAAGGAGCTCAATGCTCGCTACATGCACAAGATTGCTTCCCATCCCTACAGAGATG

FIGURE 3, page 13 of 20

9079 TTCTACTGCTCTAATAATTCCCCTAAGGAGGCAGGGGAGTGGGATTGAGGTCCCCAGA  
GAAAAGGGGAGACTTGAGAGAGACGCCTGCCCTGGCCCCACCTTAGGGCCAATCCCCATT  
TCCACTCTGGGGTTTGCAGGTGGTGGTGGCAACAGCTATCTCGGGGGGCTGTAAGATGCC  
CAAAAAGTATCACATGCAGATCGTGGGAGAAATCCAACGCGGGTGAGTCCAGGTGGCCCA  
GAAGCCTGGCCACCCGCACCTCATGCCCCACTAAGGCCTGAGCTCGGAGAGGGAGACAA  
[G,A]  
ATGAACTCTATGAAAGTGCAGTCGAAACTGTATGACACTGACCATGTATGAATTATTACT  
ATTACCGTTTCTGAGAAGGGCCGCACAACCAGCCAATGTAGGCTATTTTATGAGAAATG  
AGTCTTAACTGCCACACTCCCCTTATAAATCTCATTCAACTGATGCTGTTAAACAAAGCC  
TCTCTGAACAGCCGCTTGCTGGCTCTTTGCCTTGCTCTAATGCATTGGTTCTTTGTCCAT  
GTAGAAAGGGAATATTAGGTTCAACCAGATTTCATGAAGCATCCACTCTGTGCCAGGCAC

9537 AACTGATGCTGTTAAACAAAGCCTCTCTGAACAGCCGCTTGCTGGCTCTTTGCCTTGCTC  
TAATGCATTGGTTCTTTGTCCATGTAGAAAGGGAATATTAGGTTCAACCAGATTTCATGA  
AGCATCCACTCTGTGCCAGGCACCATGCTGGGCCCTGGGAGGAGAGGGGTGACGCTTGTC  
CTGCAGGGTTGGAACAGGCAAGGGAGGGAAGACCACATAGCACCAAAGGTCTAGGGGTCT  
GTGGACTCGTGAGCATACAGGGTTCAGAATCTGGGAGTTAACAAACGAGGCCCTACCACA  
[G,T]  
ACTGGCCCCGGGACCTTGGGCAAGTTAGGTTCTCTCAGCCTCAGTTTCTCTCTTTGTAAA  
ACAGGAGTGATGGTCCCTACCTATGGGGTGGTCTGAGGATTGAGACTGGATGGGATAA  
CTTAGGCAAAGATCCCGGCACACCATGGGGGCCCTGGCTGGTCCCTGTGGGCTGGTGAAGG  
ACTTGGCTGCCCTCCCCACTCACACCTTGGGTTCTGCCTCCTTCTGGCTCCTCGGCAG  
GTTCCCCACCCCGGTGTCGCCTGTGGTCTCACAGTGGAAGGACATGATAGGCACAGCCTT

12302 AGCCCCACCATAACCTATGGGAGAGGATTTACTAACTTTCTTAACGGTGATGAAACCAA  
GGCTCAGAATGGTTAAGTAAATTGTCAAAGGCCACAGAGGTAGGGAGTGGTAGAGTCTGG  
ATTAAACTTCCAAGTCTTGACTCCAGACCTCTAGGCTGTACTGTCTCATAGGGAAGGCA  
GTCTCACCACCTAGGGCAGAGAAGAAATCCTTAAAGCCAGAGAAGTGAGTGGCTCATC  
TGTGGTCACCCAGAGAGACAGTGATGAGGACAGGGAGAAAAATTATACCTCAGTTCACG  
[C,G]  
CCAAGGATCTGCTTTGACCATAACCCAACAAGCCCCGCTATGGTGGTATTTCTTAGGT  
TCATATGGCGGCTTTTGTTCATTGATCTTCACAGCAATTCTCTACAGGAATCTGGGC  
AGATTTATTTCTTTAGAGGAATTTCCAGGTCTTAAATCTATAGGGGGCAACTATCAAA  
ACTTCACCAATGTTGCCCCCTACCCACACACAAAACAGGCCCCAGCCGATCAGAAAG  
CACTGCTGAGCTCCTGTGAGGGCCACGCAGCTCGCTGTGAGACAGAGAGAGGGAATCA

12354 GAAACCAAGGCTCAGAATGGTTAAGTAAATTGTCAAAGGCCACAGAGGTAGGGAGTGGTA  
GAGTCTGGATTAAACTCCAAGTCTTGACTCCAGACCTCTAGGCTGTACTGTCTCATAG  
GGAAGGCAGTCTCACCACCTAGGGCAGAGAAGAAATCCTTAAAGCCAGAGAAGTGAGT  
GGCTCATCTGTGGTCACCAGAGAGACAGTGATGAGGACAGGGAGAAAAATTATACCTCA  
GTTCCAGCCCAAGGATCTGCTTTGACCATAACCCAACAAGCCCCGCTATGGTGGTATT  
[C,T]  
CCTTAGGTTTCATATGGCGGCTTTTGTTCATTGATCTTCACAGCAATTCTCTACAGGA  
ATCTGGGCAGATTTATTTCTTTAGAGGAATTTCCAGGTCTTAAATCTATAGGGGGCAA  
CTATCAAACTTCACCAATGTTGCCCCCTACCCACACACAAAACAGGCCCCAGCCGA  
TCAGAAAGCACTGCTGAGCTCCTGTGAGGGCCACGCAGCTCGCTGTGAGACAGAGAGAG  
GGAATCACATTTATTGATCACCTACTGAGCATCCATCACTAGGCTAGGACCGTCACATT

12487 ACCCACCTAGGGCAGAGAAGAAATCCTTAAAGCCAGAGAAGTGAGTGGCTCATCTGTGG  
TCACCCAGAGAGACAGTGATGAGGACAGGGAGAAAAATTATACCTCAGTTCACAGCCCAA  
GGATCTGCTTTGACCATAACCCAACAAGCCCCGCTATGGTGGTATTTCTTAGGTTTCAT  
ATGGCGGCTTTTGTTCATTGATCTTCACAGCAATTCTCTACAGGAATCTGGGCAGAT  
TTATTTCTTTAGAGGAATTTCCAGGTCTTAAATCTATAGGGGGCAACTATCAAACTT  
[C,T]  
ACCCAATGTTGCCCCCTACCCACACACAAAACAGGCCCCAGCCGATCAGAAAGCACTG  
CTGAGCTCCTGTGAGGGCCACGCAGCTCGCTGTGAGACAGAGAGAGGGAATCACATTT  
ATTGATCACCTACTGAGCATCCATCACTAGGCTAGGACCGTCACATTCTTAACTTTGA  
ATCCTTTTCATGAGGTAGGCATTATTATTCTCCTTTTGTTCACATAGCCATTAAAGAACA  
AAATTTGGGGCTGGGTGTGCTGACTCACACCTGTGATCTAGCACTTAGGGGGCTGAGGC

13198 CTAATATTAGGAAGGTTAGGCGGGAGCACAACTTGGGTTCCAGGGTTTGAAGGCTCCAG  
TGAGCTGATCTTGCCACTGCACTACAGCCTGAGCAACAGAGCAAGACCCTGTGACTCCAA  
AAACAAACAAACACATTTTGAACCCAAACAGATCTGACCCAAGATGCATGCTCTTA

FIGURE 3, page 14 of 20

TAGATGCCACCTCCCTGTGTGCTGGGGCTTCTACTAAAAACACAGACAAGATCAGGCCAAC  
CACAGTCAATCTAAGGGAAAGAGGAAAGTGTAAACAAAGCACAAATACATAAAATATTGC  
[-, A]  
AAAATGCTATTTAAAGAAAAAAAAGAGAAGAGAGGCTCTGAGGTTGTACTAACAGAGAAT  
GGCCTTGGCTAATCCAGGAAGACTTCTGAAAGAGGTTGTTTTTTCCCCAGGCTGCTTT  
TGACATCTCTCTTTTACAGTGCATCTGGGTAGTGAGCTTCTCTCTCTCTCTCTCTCA  
GCCTGCCCTATGGTGTGGCAGTGGGTGTGCCTTCTCCGTCCCTGGTCTGTGGTCTTCCAGA  
CTCAGTTGTAAGTGATAGCTTCCGCCCTCCTAGGCCACAGTCGGTTCCTGGGCCAGCC

13257 GTGAGCTGATCTTGCCACTGCACTACAGCCTGAGCAACAGAGCAAGACCCTGTGACTCCA  
AAAACAAACAAACAAACACATTTTGAACCCAAACAGATCTGACCCAAGATGCATGCTCTT  
ATAGATGCCACCTCCCTGTGTGCTGGGGCTTCTACTAAAAACACAGACAAGATCAGGCCAA  
CCACAGTCAATCTAAGGGAAAGAGGAAAGTGTAAACAAAGCACAAATACATAAAATATTG  
CAAAAATGCTATTTAAAGAAAAAAAAGAGAAGAGAGGCTCTGAGGTTGTACTAACAGAGA  
[A, G]  
TGGCCTTGGCTAATCCAGGAAGACTTCTGAAAGAGGTTGTTTTTTCCCCAGGCTGCTTT  
TTGACATCTCTCTTTTACAGTGCATCTGGGTAGTGAGCTTCTCTCTCTCTCTCTCTC  
AGCCTGCCCTATGGTGTGGCAGTGGGTGTGCCTTCTCCGTCCCTGGTCTGTGGTCTTCCAG  
ACTCAGTTGTAAAGTGATAGCTTCCGCCCTCCTAGGCCACAGGCTTCCCTGGGCCAGC  
CCGCAAAGGGCTTCCATGCCACGGCCTGGCTTAGTCCACTGTACCTTCCACCTCTGGGCC

14541 TCATGGACACTGACATTTATGTGAATCCCAAGACCTATAATAGGGTAGGTAATTCAAGCT  
TATGACCTCCTTCTTTTGGCTCTGCACCACCCCAAGAAGAGGTTGCTTTTTTAAAGCCAATA  
AAGACATTTCTGCAACTTGAGCTCAGTCTCCCTGTACAGGCCCAGGATATCCAGGGGAT  
TAAATCATCAGTACTGCTCCCCTCTCTACTTTGCCAACTCAGAGATCTTCAGGCAAAA  
GGTCATCGCCAAGGTAAGGCTCAGTCCCTGGCGACCAGAGGCTCTGGACAGAGAGTGGCC  
[G, A]  
GAAAATGGAAGCAGAAGGGCGGTGGGAGCTGAGAATAGGCCACTCCCATAGAGGGTGGAG  
GTCAAGATTGCTGTTGGCTCTCTCCCTGCAGACAGGCATGGACCCCCAGAAAGTATTACT  
AGCCAAGCAAAAATACCTCAAGAAGCAGGAGAAGCGGAGAATGAGGCCACACAACAGAG  
GAGGTCTCTATTATGAAAACCAAGGTGAATGAAGGCCAGAAGCAGCCCCGTGCCCTGCT  
CTCCTGCCCATTTCTGATACTGCCCCCTGTTACTCATGGTACCCTGGGGGCCCCGCTTCCC

14545 GGACACTGACATTTATGTGAATCCCAAGACCTATAATAGGGTAGGTAATTCAAGCTTATG  
ACCTCCTTCTTTTGGCTCTGCACCACCCCAAGAAGAGGTTGCTTTTTTAAAGCCAATAAGA  
CATTTCTGCAACTTGAGCTCAGTCTCCCTGTACAGGCCCAGGATATCCAGGGGATTTAA  
ATCATCAGTACTGCTCCCCTCTCTACTTTGCCAACTCAGAGATCTTCAGGCAAAAGGTC  
ATCGCCAAGGTAAGGCTCAGTCCCTGGCGACCAGAGGCTCTGGACAGAGAGTGCCGGAA  
[A, G]  
ATGGAAGCAGAAGGGCGGTGGGAGCTGAGAATAGGCCACTCCCATAGAGGGTGGAGGTCA  
AGATTGCTGTTGGCTCTCTCCCTGCAGACAGGCATGGACCCCCAGAAAGTATTACTAGCC  
AAGCAAAAATACCTCAAGAAGCAGGAGAAGCGGAGAATGAGGCCACACAACAGAGGAGG  
TCTCTATTATGAAAACCAAGGTGAATGAAGGCCAGAAGCAGCCCCGTGCCCTGCTCTCC  
TGCCCATTTCTGATACTGCCCCCTGTTACTCATGGTACCCTGGGGGCCCCGCTTCCCACCC

15041 ACCAAGGTGAATGAAGGCCAGAAGCAGCCCCGTGCCCTGCTCTCTGCCCATTCTGATAC  
TGCCCCCTGTTACTCATGGTACCCTGGGGGCCCCGCTTCCCACCCTGACAGGCAAGACA  
GAAAGTCTCTGGGAACACTGCTTGGTGGCCGCTGGGCATTTTTTCTTCTTTTTTTCTTTT  
TCTTTTTAGAGATGGAATTTTGTCTTGTACCCAGGCTTGAGTGCAATGGCGTTATCTT  
GGTCACTGCAACCTCCACCTCTGGGGTTCAAGCGATTCTCCTGCCTTAGCCTCCCAAGT  
[C, A]  
GCTGAGATTACAGGTGCCACCACACCCAGCTAATTTTTTGTATTTTTTAGTAGATATTGGGT  
TTACCATGTTGGCCAGGCTGGTGTCAAACCTCTGACCTCAGGTGATCCACCTACCTTAG  
CCTTCCAAAGTGCTGGGATTACAAGCCTGAGCCACTGCGCCAGCCTGGGCATTTTTCTT  
CTTGGATGAGGTGCTACCATCTCCCAGGGAAGCCACTGAACCCCCAAGGCCCTTCTCCAT  
TTTCTGGCTAAGATAGGACATGGCCCATGGACTTTTGAACAACCCAGAGGGGGAACAGCA

15053 GAAGGCCAGAAGCAGCCCCGTGCCCTGCTCTCCTGCCCATTTCTGATACTGCCCCCTGTTA  
CTCATGGTACCCTGGGGGCCCCGCTTCCCACCCTGACAGGCAAGACAGAAAGTCTCTGG  
GAACACTGCCTGGTGGCCGCTGGGCATTTTTTCTTCTTTTTTTTCTTTTTCTTTTAGAGA  
TGGAATTTTGTCTTGTACCCAGGCTTGAGTGCAATGGCGTTATCTTGGCTCACTGCAA  
CCTCCACCTCTGGGGTTCAAGCGATTCTCCTGCCTTAGCCTCCCAAGTCGCTGAGATTAC  
[A, C]

GGTGCCACCACACCCAGCTAATTTTTGTATTTTTAGTAGATATTGGGTTTCACCATGTTG  
GCCAGGCTGGTGTCAAACCTCCTGACCTCAGGTGATCCACCTACCTTAGCCTTCCAAAGTG  
CTGGGATTACAAGCCTGAGCCACTGCGCCAGCCTGGGCATTTTCTTCTTGATGAGGT  
GCTACCATCTCCAGGGAAGCCACTGAACCCCCAAGGCCCTTCTCCATTTTCTGGCTAAG  
ATAGGACATGGCCCATGGACTTTTGAACAACCCAGAGGGGGAACAGCAGTGAATTCCTG

15065 CAGCCCCGTGCCCTGCTCTCCTGCCATTCTGATACTGCCCCCTGTTACTCATGGTACCC  
TGGGGGCCCCGCTTCCCACCCTGACAGGCAAAGACAGAAAGTCTCTGGGAACACTGCCTG  
GTGGCCGCTGGGCATTTTCTTCTTTTTTCTTTTTCTTTTTAGAGATGGAATTTTGCT  
CTTGTCACCCAGGCTTGAGTGCAATGGCGTTATCTTGGCTCACTGCAACCTCCACCTCTG  
GGGTTCAAGCGATTCTCCTGCCTTAGCCTCCCAAGTCGCTGAGATTACAGGTGCCACCAC  
[A, G]  
CCCAGCTAATTTTTGTATTTTTAGTAGATATTGGGTTTCACCATGTTGGCCAGGCTGGTG  
TCAAACCTCCTGACCTCAGGTGATCCACCTACCTTAGCCTTCCAAAGTGCTGGGATTACAA  
GCCTGAGCCACTGCGCCAGCCTGGGCATTTTCTTCTTGATGAGGTGCTACCATCTCC  
CAGGGAAGCCACTGAACCCCCAAGGCCCTTCTCCATTTTCTGGCTAAGATAGGACATGGC  
CCATGGACTTTTGAACAACCCAGAGGGGGAACAGCAGTGAATTCCTGGGGAACCCAGGC

15108 TGTACTCATGGTACCCTGGGGGCCCCGCTTCCCACCCTGACAGGCAAAGACAGAAAGTC  
TCTGGGAACACTGCCTGGTGGCCGCTGGGCATTTTCTTCTTTTTTCTTTTTCTTTTT  
AGAGATGGAATTTTGCTCTTGTCACCCAGGCTTGAGTGCAATGGCGTTATCTTGGCTCAC  
TGCAACCTCCACCTCTGGGGTTCAAGCGATTCTCCTGCCTTAGCCTCCCAAGTCGCTGAG  
ATTACAGGTGCCACCACACCCAGCTAATTTTTGTATTTTTAGTAGATATTGGGTTTCACC  
[A, C]  
TGTGGCCAGGCTGGTGTCAAACCTCCTGACCTCAGGTGATCCACCTACCTTAGCCTTCCA  
AAGTGCTGGGATTACAAGCCTGAGCCACTGCGCCAGCCTGGGCATTTTCTTCTTGAT  
GAGGTGCTACCATCTCCAGGGAAGCCACTGAACCCCCAAGGCCCTTCTCCATTTTCTGG  
CTAAGATAGGACATGGCCCATGGACTTTTGAACAACCCAGAGGGGGAACAGCAGTGAATT  
TCCTGGGGAACCCAGGCAGCCAGGGCTAGCAAGGCTGGGGTGGCCATGGCAGTAATCCT

16274 CTTCCAGACTGTCTCCCTGCAGGAGCTGCAGCAGGACTTTGAGAATGCGCCCCCACCGA  
CCCCAACAAACACAGACCCCGGCTAACGGCACCAGCGTGTCTATATCACCTTCAGCCC  
TGACAGCTCCTCACCTGCCCAGAGTGAGCCACCAGCCTCCGCTGAGGCCCCCGCGAGCC  
CAGTGACATGCTGGCCAGCGTCCCACCCTTCGTACCTTCCACACCCTCATCCTGGACAT  
GAGTGGAGTCAGCTTCGTGGACTTGATGGGCATCAAGGCCCTGGCCAAGGTGAGGCCCTC  
[-, G]  
GGGACAGCAAGCACCACCCACTCCACCCCTCCGCTCTGCTCTCCACATTCCCTTTCTG  
GGAGCCCTCATTTTCAGGAAGCTGAGGGAGGAAGCTCACTGGGGAGACTAACAGCTCCTAG  
GAATCCCTCCTTTCCCAGACGCCACCAGGTTGAGACATTCTCCACAGAGCAGGCCCAGA  
CGGCCCATGACAATGAGTGGCGGGACAAGTCTACCAGAGTTTCAGGCCCTGTGCTCCCA  
ACACCCAGCAGTGGCCATCCCAAGTCCCTCTCAGCCATCAGGAACCCACCCAGGTTCT

17424 AACATGGTGAAACCCCGCTCTACTAAAAATACAAAAATTAGCCAGGTGTGGTGACGGGC  
CCCTGTAGTCCCAGCTACTCGGTAGGCTGAGGCAGAGAATTGCTTGAACCCAGGAGGCGG  
AGGTTGCAGTGAGCCAAGATCGCGCCACTGCACTCCAGCCTGGGCAACAGAGTGAAACTC  
CATCTCAAAGAAAAAAGAAAAATATCTAGCCCCACAAGAAGGGGCCATGGTGACTTT  
AAGTGCCCGCCACGTTGGCAAAAGTCCATTTCCGCTCCACTTCCCAGAGAAACCGTCAGC  
[C, T]  
AACACTCCAGGGAGAAGTGGTGTGCTTTGCTGCTATTTTGTCTTTGGCTGCTGGGCTCT  
CAGGGTTGCTTATTTGTTGGCTTCCCTCTGAAGTACGTTTTGTGAATCACTTTTGAGA  
CCCACTCAGAACATTCTTTCTTTTGCTCCCTACCCCAACAACACTTCTAGCTGAGCT  
CCACCTATGGGAAGATCGGCGTGAAGGTCTTCTTGGTGAACATCCATGGTAAGAGAAAGA  
GGACATTTAGGGACTGAAAGACTGGCAAGGAGTGTGGGGTAGGAACAGGTTGGTGGGGTC

17627 AATATCTAGCCCCACAAGAAGGGGCCATGGTGACTTTAAGTGCCCGCCACGTTGGCAAAA  
GTCCATTTCCGCTCCACTTCCCAGAGAAACCGTCAGCCAACACTCCAGGGAGAAGTGGTG  
TGCTTTGCTGCTATTTTGTCTTTGGCTGCTGGGCTCTCAGGGTTGCTTATTTGTTTGGC  
TTCCCTCTGAAGTACGTTTTGTGAATCACTTTTGAGACCCACTCAGAACATTCTTTCC  
TTTTGCCCTCCCTACCCCAACAACACTTCTAGCTGAGCTCCACCTATGGGAAGATCGGCGT  
[G, A]  
AAGGTCTTCTTGGTGAACATCCATGGTAAGAGAAAGAGGACATTTAGGGACTGAAAGACT  
GGCAAGGAGTGTGGGGTAGGAACAGGTTGGTGGGGTCTGAATAGTGAGGAGGTTGGAAC  
GAGAGACCCAGCTATCCCCACAAGCTGCTGCCTGCTCATAAAGCTTCAGGTACAAGT

FIGURE 3, page 16 of 20

CCAAAGAGACTGGTCAGATTGCATAAACATCCTAGGGGCCCTAGTGACAGAGTGGGGGTG  
AGGAGGTCATGGAGTTACAGAAGGACAGCTAGGATTCTAATCTACCCATAACTAATTTG

18427 GGGTGCATATACACAGCCTCAAGGACGTGGCCACAGGGCAGCAGACATTTACATGACTAG  
CATGTACGCAAAAGTGCAGAGATGTGGGAGCAAGTGCACACAGACACACAGGAGAATGTGA  
AGGGGCACATACACACACACCCAGCTCCCTGCACTGGGTGAGACCCCTCCAGCAGGGCT  
GCAGTTCCCAAGCTCCGCATGGCCACGTTCGGGGAGAGAATCTGCAGTGGCAATGACCTG  
CTATGATATGTTCTGGAGTTAGAAGCAGTGGATTCTCCCCAACCTCACTGGACACCCCT  
[T, C]  
AGGAAACCATCTCTAGGATTAAGAGTAATCCACACAACTTCCAATGCCACACATTGGAA  
GTTGCTGGAAGGTCTGGGAAAACAAGAGGAAGGATGGGTCTTGGGGGATAGAACTGGC  
AGCGGCTCTTCAAGGATGGCTTAGGCTTTTCCACTCGAATCACCACAAAGTACTGACTC  
CCTAAATCAAAGTCTCTCTGCTCTGGGTGAACTTCAGCATCCTCAAGTTCATGT  
TGCCCTCTGCCGTCCAGAACTGATATTGCACTGCCAATGCCATGGCCCTCAGATACAGCA

18813 AGAGGAAGGATGGGTCTTGGGGGATAGAACTGGCAGCGGCCCTCTCAAGGATGGCTTAG  
GCTTTTCCACTCGAATCACCACAAAGTACTGACTCCCTAAATCAAAGTCTCTCTGCTG  
TCTGGGTGAAACTTCAGCATCCTCAAGTTCATGTTGCCCTCTGCCGTCCAGAACTGATA  
TTGCACTGCCAATGCCATGGCCCTCAGATACAGCAAGAGCTGGGACCTCAGGCCCTCTCC  
ATCCCTGCTCTGGTCTCACTATCTTCCCCACCCCACTCCAATCCACAATGGCTGTTAT  
[C, G]  
TTTCTGAAGGTGATCTTTTCTCCTTCTAGCCAGGTGTACAATGACATTAGCCATGGAGG  
CGCTTTTGGAGTGGGAGTCTAGAATGCAAGCACGTCTTCCAGCATACATGACGCAGT  
CCTCTTTGCCAGGCAAAATGCTAGAGACGTGACCCAGGACACAACTTCCAAGGGGTAAAG  
GTTCTTGCACTGGGGAATCCTAGGCTCCAAGGCACTGAAATAGCAGGACCAAGAGGCAT  
TATTAGAAAGAACACAGGAGAAGGTTTAAAGTTCCAATATCAAGTCTGCCATTTCAAGTTT

19035 GGACCTCAGGCCTCTCCCATCCCTGCTCTGGTCTCACTATCTTCCCCACCCCACTCCA  
ATCCACAATGGCTGTTATCTTTCTGAAGGTGATCTTTTCTCCTTCTAGCCAGGTGTACA  
ATGACATTAGCCATGGAGGCGTCTTTGAGGATGGGAGTCTAGAATGCAAGCACGTCTTTT  
CCAGCATACATGACGCAGTCTCTTTGCCAGGCAATGCTAGAGACGTGACCCAGGAC  
ACAATCTCCAAGGGGTAAAGTCTTGACCTGGGGAATCCTAGGCTCCAAGGCACTGAAA  
[T, C]  
AGCAGGACCAAGAGGCATTATTAGAAAGAACACAGGAGAAGGTTTAAAGTTCCAATATCAA  
GTCTGCCATTTCAAGTTTCTGAATCTGTTTCTTATCTATAGAATGAGCACCATCAACTA  
ACATTACCTACCTCTCTGCATTTTCTTTTATTTTGTGTTTAGGGTTAAATGATAATTACA  
TCTTTTGTGTCACTTGAAGGCTTTTGTGATTGTAAAAATCTTTATCAATATAAGTTT  
TCTGGTTGCACAAACACCCAAAGCATAGTAGAGCAGGCCACTCTGCTGGCATCGTTCCC

19182 AGGATGGGAGTCTAGAATGCAAGCACGTCTTTCCAGCATACATGACGCAGTCTCTTTG  
CCCAGGCAATGCTAGAGACGTGACCCAGGACACAACTTCCAAGGGGTAAAGTTCTTGC  
ACCTGGGGAATCCTAGGCTCCAAGGCACTGAAATAGCAGGACCAAGAGGCATTATTAGAA  
AGAACACAGGAGAAGGTTTAAAGTTCCAATATCAAGTCTGCCATTTCAAGTTTCTGAATCT  
GTTTCTTATCTATAGAATGAGCACCATCAACTAACATTACCTACCTCTCTGCATTTTTT  
[T, C]  
TTTATTTTGTGTTTAGGGTTAAATGATAATTACATCTTTTGTGTCACTTGAAAGCACTTTG  
TGTATTGTAAAAATCTTTATCAATATAAGTTTTCTGGTTGCACAAACACCCAAAGCATA  
GTAGAGCAGGCCACTCTGCTGGCATCGTTCCCTGCCTCCTCCTCATCTCTTTCTAAAGG  
GGGCTTTTCGGGAAGGGAGGGGAGGGGAGTAAGCCTACCCATTTTAACTTACCGGAGCTTA  
GAGATTTCAAGCTGGTGAGGGATAAAGAGATTGGGTCTGAGTTTGTCTCAGCTTTTTGA

19508 TAATTACATCTTTTGTGTCACTTGAAAGCACTTTGTGTATTGTAAAAATCTTTATCAAT  
ATAAGTTTTCTGGTTGCACAAACACCCAAAGCATAGTAGAGCAGGCCACTCTGCTGGCA  
TCGTTCCCTGCCTCCTCCTCATCTCTTTCTAAAGGGGGCTTTTCGGGAAGGGAGGGGAGGG  
GAGTAAGCCTACCCATTTTAACTTACCGGAGCTTAGAGATTTCAAGGCTGGTGAGGGATAA  
AGAGATTGGGTCTGAGTTTGTCTCAGCTTTTTGACATTTAATTTACTAGCTCAGTAAGT  
[-, G, C]  
ATACAAATGGGATACAAATAACACCATCTAAACTCCAGAAGACTGGGGAGTCAGAAAAA  
TCCTACCTCCTTGGGGTCCCTGCCAGATCCCAAGTCTCTAGCCCTCAGGGTCCCCT  
CCCAGCTCAGCTCCTGCCCTTGGCCCTCCCAAGACTCTTGTGTGCCCCAGCCCTGGGTAA  
AAACCTCCCCTGCCCTCTGTGGGTGATAAGAAAGGCTTTTCTGGCCCTAGAGCAATGATT  
TGCTCTTTGCCCTTAAAGAGACTGATGAAGGTGAAACCATCTGTTCTAAGTGTGAAAGACT

FIGURE 3, page 17 of 20

008321 "6855h460

19571 AGTTTTCTGGTTGCACAAACACCCAAAGCATAGTAGAGCAGGCCCACTCTGCTGGCATCG  
TTCCCTGCCCTCCTCCTCATCTCTTTCTAAAGGGGGCTTTCGGGAAGGGAGGGGAGGGGAG  
TAAGCCTACCCATTTTAACTTACCGGAGCTTAGAGATTTACGGCTGGTGAGGGATAAAGA  
GATTGGGTCTGAGTTTTGTCTCAGCTTTTGTACATTTAATTTACTAGCTCAGTAAGTCAT  
ACAAATGGGATACAAATAACACCATCTAAACTCCAGAAGACTGGGGAGTCAGAAAAATC  
[T, G, C]  
TACCTCCTTGGGGTCCCTGCCAGATCCCCAGTCATCTCTAGCCCTCAGGGTCCCTCCC  
AGCTCAGCTCCTGCCCTTGGCCTCCCAAGACTCTTGTGTGCCCCAGCCCTGGGTAAAAA  
CCTCCCTGCCCTCTGTGGGTCTAAGAAAGGCTTTTCTGGCCCTAGAGCAATGATTTGC  
TCTTTGCCCTTAAGAGACTGATGAAGGTGAAACCATCTGTTCTAAGTGCTGAAAGACTGCC  
CAGGAACACACAGGGCGCTGGCTCCTGCCCTCCATGCCTAGAGGGAAACCTGGGGAAAC

20147 GCCTAGAGGGAAACCTGGGGAAACAACGGGCTTTCCTGCTTCGTGAAATTTGTCCGCAG  
AGCAAAGAGGGAGATTCTGGAGGAAGCTGCATTAGTTGTTAGTGCCCTAATCATGTTTCAG  
CTACTCTAGTTGGTATGTATACTTGATTAGTCATAGCACTTATAAATAATTTATATTTTA  
TATAATATATACTTACATATTATAGACCATTACAGATACAAATCACACACATAAACACA  
CACCTTTTCAACAGCATTGTGAGGGACAAAGCAGGCAAAGTGAGGCTGGTTATCAGACTT  
[T, G]  
AACAGATTAGAAAATATATTCCCAGGAGGACAGGAATTCCTCAAGGTGAGGCAGCTAGCC  
AATAGTTTTTCTAAGCTGAGTAAAACCTTCCCTGCCCTCTAACGGCCCAAAAGGAGGGAA  
GACCGCGATACACACCTGTCTGGTATAAGGGGGAAGACCACAGCCGTGCTGTTTTGTGA  
GGCAGGTAAGGGGAAGGGGCAAGAGGATAAGTCATGTGTGAGGAAGCAGCGTCCAACCAGA  
GCCGGCCACCTGTCCCTTTTCTGCCACCATGCACCAACTTTGCTGTTTCACTCACTGAAG

20180 TTCCTGCTTCGTGAAATTTGTCCGCAGAGCAAAGAGGGAGATTCTGGAGGAAGCTGCATT  
AGTTGTTAGTGCCCTAATCATGTTTCAGCTACTCTAGTTGGTATGTATACTTGATTAGTCA  
TAGCACTTATAAATAATTTATATTTTATATAATATATACTTACATATTATAGACCATTCA  
CAGATACAAATCACACACATAAACACACACCTTTTCAACAGCATTGTGAGGGACAAAGCA  
GGCAAAGTGAGGCTGGTTATCAGACTTTAACAGATTAGAAAATATATTCCCAGGAGGACA  
[G, A]  
GAATTCCTCAAGGTGAGGCAGCTAGCCAATAGTTTTTCTAAGCTGAGTAAAACCTTCCCT  
GCCTCTAACGGCCCAAAAGGAGGGGAAGACCGCGATACACACCTGTCTGGTATAAGGGGG  
AAGACCACAGCCGTGCTGTTTTTGTGAGGCAGGTAAGGGGAAGGGGCAAGAGGATAAGTCA  
TGTGTGAGGAAGCAGCGTCCAACCAGAGCCGGCCACCTGTCCCTTTTCTGCCACCATGC  
ACCAACTTTGCTGTTTCACTCACTGAAGCTCATTCTGCACTGGCTTCCCTCCCTCCAGGCT

20584 TGTCTGGTATAAGGGGGAAGACCACAGCCGTGCTGTTTTTGTGAGGCAGGTAAGGGAAGG  
GGCAAGAGGATAAGTCATGTGTGAGGAAGCAGCGTCCAACCAGAGCCGGCCACCTGTCCC  
TTTTCTGCCACCATGCACCAACTTTGCTGTTTCACTCACTGAAGCTCATTCTGCACTGGC  
TTCCTCCCTTCCAGGCTCCAGGGGATGCTGAGCTCTCCTTGTACGACTCAGAGGAGGACA  
TTCGCAGCTACTGGGACTTAGAGCAGGTGAGCTGAGGGAAGGGGCTGTGAGGGTGGGAGC  
[A, T]  
GGGCGAAGAGGGGAAGGATGGGGTCTGTCAAATACAAGGCGTTCACTCAGCTGTCTCA  
CCTCCAGCCCAAGAGCAGTCACATTCAGGCCACAAAGATTTGTGGTCATCTTTGTTTTTT  
TTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT  
AGACTGGAATGCAGTGGCATGATCTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCCAGA  
GGTTCTCCTGCCTCAGCTCCCAGTAGCTGGGACTTCAGGCTGCGCCAGCTAATTTT

20717 ATGCACCAACTTTGCTGTTTCACTCACTGAAGCTCATTCTGCACTGGCTTCCCTCCCTTCCA  
GGCTCCAGGGGATGCTGAGCTCTCCTTGTACGACTCAGAGGAGGACATTTCGAGCTACTG  
GGACTTAGAGCAGGTGAGCTGAGGGAAGGGGCTGTGAGGGTGGGAGCAGGGCGAAGAGGG  
GAAGGATGGGGTCTGTCAAATACAAGGCGTTCACTCAGCTGTCTCACCTCCAGCCAG  
AGCAGTCACATTCAAGGCCACAAAGATTTGTGGTCATCTTTGTTTTTTTTCTTTTCTTTT  
[T, C]  
CTTTTTTTTTTTTTTTTAAATTTGAGACAAAGTCTCACTCTATCACCCAGACTGGAATGCA  
GTGGCATGATCTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCCAGAGGTTCTCCTGCCT  
CAGCCTCCCGAGTAGCTGGGACTTCAGGCCTGCGCCAGCTAATTTTTGTATTTTTTAGTA  
GAGACAGCTTTTACCATGTTGGCTGGGCTGGTCTCGAACTTCCGATCTCAAGCAATCTG  
CCTGCCTCGGTCTCCTAAGTGCTGGATTACAGGCATAAGCCACGATGCCTGGCCTTTGT

20894 GGGGAAGGATGGGGTCTGCTGTCAAATACAAGGCGTTCACTCAGCTGTCTCACCTCCAGCC  
CAGAGCAGTCACATTCAAGGCCACAAAGATTTGTGGTCATCTTTGTTTTTTTTCTTTTCC  
TTTTCTTTTTTTTTTTTTTTTAAATTTGAGACAAAGTCTCACTCTATCACCCAGACTGGAA

FIGURE 3, page 18 of 20



TGCAGTGGCATGATCTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAGAGGTTCTCCT  
GCCTCAGCCTCCCGAGTAGCTGGGACTTCAGGCCTGCGCCAGCTAATTTTTGTATTTTT  
[A, G]  
GTAGAGACAGCTTTTACCATGTTGGCTGGGCTGGTCTCGAACTTCGATCTCAAGCAAT  
CTGCCTGCCTCGGTCTCCTAAGTGCCTGGATTACAGGCATAAGCCACGATGCCTGGCCTT  
TGTTTTCACTCTCTCACTCCCTGAAAGGCATCGTGGGAGAGGGTGAGTCACTGGACCA  
AGTCCTAGAGAACCAGTATCTATTCTTATTCTCCAACACATCACCACGTCACCTGAGC  
AAGCCACATACACCCTGGGCCCTAGTTTTTATCATCTGTGAAATTAGGGGAAACATAGGT

21787 GGGTGCAATGGTTACACCTGTAATCCCAGCACTTTGGGAGGCTGAGGTGGGCGGACCAC  
CTGAGGTGAGGAGTTTGGAGCCAGCCTGGCCAACATGGCGAAACCCCGTCTCTACTAAAA  
GCACAAAAATTAGCCAGGCGTAGTGGTGCATGCCTGTAGTCCCAGCTACTCGGAAGTCTG  
AGGCATGAGAATCACTTGAACCTGGGAGGCAGATGTTGCAGTGAGCCGAGATCGTGCCAC  
TGCACTCCAGCTTGGGTGACAGAGCTAGACTGTCTCAAAAACAAACAAACAAACAAAC  
[-, A, C]  
TAAAAGATATGTGGATATGAGGGATCACCATCCCCATAGGGCCCCCTGGATTAACACCACC  
CCACCAATGCCCTGAATTAAAAGAAACCAGATGACTAGGTTTGGAGAAATCTGGCTTTGG  
GTCTATGAGAAGTAGTGTCTCTCTTTGTGCCTCTTCCCATTCTTTTTGACATTGAGCTCC  
ATGGTGCTCTGAATCCGTCTCTCACAGTGTGATGGCAGGTGGGACAGATTAGAAAATAG  
AGCTGGAGCCACAGAGATTTGGCAGACTGATTTCCGGTGCCTCTTGAATCTCCAGCACA

22264 CTCCATGGTGCTCTGAATCCGTCTCTCACAGTGCTGATGGCAGGTGGGACAGATTAGAAA  
ATAGAGCTGGAGCCACAGAGATTTGGCAGACTGATTTCCGGTGCCTCTTGAATCTCCAG  
CACATTTCAAAAAGCCTGGATAGGACCAAAATAGCTTATCAACGTGAGAAAGGACTTCAG  
AGCTTGTCTACTGCCAACCTCATTTTACCCAATGAGGAAAGTGAAGCTATTAGGGGGCG  
AGGGACACGTGGAAGGTACACAGCACACAGGAGGTGATTCACATGTAGATTTAGCACC  
[T, C]  
GCTCCTGCCACGCTGGACTGGTTACCTCCTAGGCTGACCTGCCTCTCCCCTGTTTACA  
CACACTCTCGCACACACACACACACACACACACAGGTGCTTTGTCTGGCCAGG  
GGTTCTAGGGTCACTCTTTGGTTGCAGCCACTGTGACCCAACTGGTCTAACCTCTCTC  
TTCCCCTCCCACTTCTTCTGTGGTTCTGTCAGGAGATGTTGGGAGCATGTTTACGC  
AGAGACCCTGACCGCCTGTGAGGGCTCAGCCAGTCTCATGCTGCCTACAGAGTGCCTG

22338 ACAGAGATTTGGCAGACTGATTTCCGGTGCCTCTTGAATCTCCAGCACATTTCAAAAAG  
CCTGGATAGGACCAAAATAGCTTATCAACGTGAGAAAGGACTTCAGAGCTTGTCTACTGC  
CAACCCTCATTTTACCCAATGAGGAAAGTGAAGCTATTAGGGGGCGAGGGACACGTGGAA  
GGTCACACAGCACACAGGAGGTGATTCACATGTAGATTTAGCAGCCTGCTCCTGCCACGC  
TGGACTGGTTTACCTCCTAGGCTGACCTGCCTCTCCCCTGTTTACACACACTCTCGCAC  
[-, C, A]  
CACACACACACACACACACACACAGGTGCTTTGTTCTGGCCAGGGGTTCTAGGGTCA  
CCTCTTGGTTGCAGCCACTGTGACCCAACTGGTCTAACCTCTCTCTTCCCCTCCCACTT  
CCTTCTGTGGTTCTTGCAGGAGATGTTCCGGGAGCATGTTTACGCAGAGACCCTGACCG  
CCCTGTGAGGGCTCAGCCAGTCTCATGCTGCCTACAGAGTGCCTGGCACTTGGGACTTC  
CATAAAGGATGAGCCTGGGGTACAGGGGGTGTGCGGCGGAGGAAAGTGCATCCCCCAGA

23363 CAGGGACCATGTGCTCTCCACACCCAGGAGTCTAGGCCTTGGTAACTATGCGCCCCCGT  
CCATCATCCCCAAGGCTGCCCAAACCACTGCTGTGAGCAAGCACATCAGACTCTAGC  
CTGGACAGTGGCCAGGACCGTCGAGACCACAGAGTACCTCCCCGGGGACAGCCACTA  
AGGTTCTGCCTCAGCCTCCTGAAACATCACTGCCCTCAGAGGCTGCTCCCTTCCCTGGA  
GGCTGGCTAGAAACCCCAAGAGGGGGATGGGTAGCTGGCAGAATCATCTGGCATCCTAG  
[T, C]  
AATAGATACCACTTATTCTGCACAAAACCTTTGGGAATTCCTCTTTGCACCCAGAGACTC  
AGAGGGGAAGAGGGTGTAGTACCAACACAGGGAAAACGGATGGGACCTGGGCCAGACA  
GTCCCCCTTGACCCAGGGCCCATCAGGGAAATGCCTCCCTTTGGTAAATCTGCCTTATC  
CTTCTTTACCTGGCAAAGAGCCAATCATGTTAACTCTTCTTATCAGCCTGTGGCCAGAG  
GACACAATGGGGTCTTCTGTAGGCAAAGGTGGAAGTCTCCAGGGATCCGCTACATCCC

23688 AAACTTTTGGGAATTCCTCTTTGCACCCAGAGACTCAGAGGGGAAGAGGGTGTAGTACC  
AACACAGGGAAAACGGATGGGACCTGGGCCAGACAGTCCCCCTTGACCCAGGGCCCAT  
CAGGGAAATGCCTCCCTTTGGTAAATCTGCCTTATCCTTCTTTACCTGGCAAAGAGCCAA  
TCATGTTAACTCTTCTTATCAGCCTGTGGCCAGAGACACAATGGGGTCTTCTGTAGG  
CAAAGGTGGAAGTCTCCAGGGATCCGCTACATCCCCTAACTGCATGCAGATGTGGAAAG  
[G, A]

FIGURE 3, page 19 of 20

GGCTGATCCAGATTGGGTCTTCCTGCACAGGAAGACTCTTTAACACCCTTAGGACCTCAG  
GCCATCTTCTCCTATGAAGATGAAAATAGGGGTTAAGTTTTCCATATGTACAAGGAGGTA  
TTGAGAGGAACCCTACTGTTGACTTGAAAATAAATAGGTTCCATGTGTAAGTGTGTTTGT  
AAATTTCACTGGAAATGCACAGAAAATCTTCTGGCCTCTCATCACTGCTTTTCTCAAGCT  
TCTTCAGCTTAACAACCCCTTCCCTAACAGGTTGGGCTGGCCAGCCTAGGAAAACATCC

24210

TCACTGCTTTTCTCAAGCTTCTTCAGCTTAACAACCCCTTCCCTAACAGGTTGGGCTGGC  
CCAGCCTAGGAAAACATCCCCATTTCTAACTTCAGCCAGACCTGCGTTGTGTGTCTGTGT  
GTTGAGTGAGCTGGTCAGCTAACAAGTCTTCTTAGAGTTAAAGGAGGGGGTGTGGCCAA  
GAGCCAACACATTCTTGGCCAGGAGCATTGCTTTTCTGTGAATTCATTATGCCATCTGG  
CTGCCAATGGAACTCAAACTTGAAGGCGAAGGACAATGTTATCTGGGATTCACCGTGC  
[A, C]  
CAGCACCCGAAGTGCCAAATTCAGGAGGACAAGAGCCTTAGCCAATGACAACTCACTCT  
CCCCTACTCCACCTCCTTCCAAGTCCAGCTCAGGCCAGGAGGTGGGAGAAGGTCACAGA  
GCCTCAGGAATTTCCAAGTCAGAGTCCCCTTTGAACCAAGTATCTAGATCCCCTGAGGAC  
TTGATGAAGTGATCCTTAACCCCCAAGTAATCATTAAACCCCAAGACCTCAGAAGTG  
AAGGAGATTGTTGACCCAGTGACCTGGAGTTGAGGCTCAGGGAGAGATCTGCCACATGTC

6356450